Natural Time and Crash Risk

Ata Türkoğlu

A thesis submitted in partial fulfillment

of the requirements for the degree of

PhD. in Computational Finance

Centre for Computational Finance and Economic Agents

University of Essex

December 2015

Table of Contents

Acknowledgments

I gratefully acknowledge the support and generosity of ETOR Charity Foundation for making this PhD. possible.

I would also like to express my deepest gratitude to my supervisor Dr. Wing Lon Ng for his invaluable comments without which this research could not have been done to its fullest.

Last but not least, I would like to thank my family - my parents Vural and Vahide; and my brother Türker, for their unwavering support and belief in me.

Publications

Chapter 2 of this thesis is forthcoming in *Quantitative Finance*.[1]

[1] The online article "Normally distributed high-frequency returns: a subordination approach" can be accessed at:

http://www.tandfonline.com/doi/abs/10.1080/14697688.2015.1023335?journalCode=rquf20.

Abstract

The deviation of financial returns from normal distribution is a well-documented stylized fact. Nonetheless, finance professionals and investors alike pay attention to these deviations almost only when a crisis erases years' worth of gains. And despite decades' worth of literature, the culprit for non-normal distribution of financial returns is still not determined with certainty. In this research, I address the non-normality of return distributions and financial crashes together. Specifically, I aim to identify the determinants of non-normality in a high frequency setting and utilize these variables to forecast financial crashes. To this effect, multiple instruments and time horizons are considered.

The contribution of this thesis is multifold. The "natural time" approach introduced here, uses order book variables to achieve normally distributed high frequency returns via subordination. In its essence, natural time is a two-step procedure which uses high frequency order book variables as a gauge for variance while sampling in transaction time. Natural time provides the reader with a new lens to view the financial markets and underscores two important aspects of the high frequency world; sampling frequency affects the distributions we observe and order book variables such as liquidity are the key to heteroscedasticity in asset returns. So much so that subordination with order book variables under transaction time achieves the normal return distribution which underlies numerous financial theories we use today.

I further extend the use of these order book variables by introducing the "market heat" metric. Market heat generates successful binary flash crash predictions and its success

adds support to the claim that liquidity concerns may be the primary driver of price formation processes. Finally, I expand the findings of this research on high frequency asset returns to a macroeconomic setting by producing currency devaluation predictions for G10 currencies. The early warning systems produced here demonstrate that not only debt related macroeconomic variables but also liquidity related market variables are at play when it comes to currency fluctuations.

Chapter 1: Introduction

1.1 Motivation

In finance we often make simplifying assumptions to describe the observable data; the cardinal assumption being normally distributed asset returns. However, empirical data, especially high frequency returns, often diverge from normal distribution substantially. As a result, inferences made using the normal distribution assumption become unreliable. And not so infrequently, we are reminded of this fact when financial crashes wreak havoc on unsuspecting investors. These "tail" events constitute a problem both for finance theory and the sustainable growth of economies. Therefore, we need to determine what makes asset returns behave so erratically.

Is it our choice of sampling frequency that changes an otherwise well behaved series? Or are there factors that we are not accounting for that might be affecting the return distributions? If so, can we use these factors to predict the next tulip mania or the Flash Crash? I posit that the answer to all these questions is a "Yes". Then using both high and low frequency datasets, I put to test each one of these assumptions.

I begin by identifying the main contributors to the non-normality of asset returns. Existing research on asset distributions primarily use calendar time sampling despite solid evidence that this type of sampling causes distortions in the data, especially in high frequency settings. Furthermore, the market conditions in which financial crises form are often neglected. This is particularly important as financial crises cause the very deviations that undermine the normal distribution assumption. The natural time approach introduced in this thesis aims to remedy these shortcomings in the extant literature by sampling in transaction time and subordinating with respect to order book variables that capture market conditions. Natural time pinpoints the elements that cause non-normality of asset returns and uses them to recover the normal distribution, all the while keeping the data intact from errors due to calendar sampling.

Natural time accounts for the heteroscedasticity in returns using contemporary order book variables. The next logical step then is to test if these variables can also be used to predict high frequency crashes. The Flash Crash of May 6^{th}, 2010 is an especially good example to study given its recency and the haste with which algorithmic traders were blamed for it. However, no matter who is to blame for the Flash Crash, one fact remains: the market was caught off guard. Hence, in this thesis, instead of looking for a culprit for the Flash Crash, I aim to create a warning system that can predict impending flash crashes both for indices and single stocks. In other words, the market heat metric introduced in this research is a potential circuit breaker that tracks liquidity conditions in the market and warns about potential liquidity driven price dips so that the stock exchanges may halt the markets to give them time to recover the much needed liquidity.

The success of market heat in predicting such episodes proves the predictive power of liquidity based order book variables. However, market heat's ability to outperform alternative warning systems may partly be attributed to perfect classification of trades. Existing flash crash literature is profuse with methods that classify trades in bulk, introducing errors into the original data. Although not one of the explicit objectives of this research, true classification of trades into bid or ask initiated transactions is achieved at all times in this thesis. Hence, both market heat and natural time produce accurate inferences about high frequency asset returns.

A high frequency episode like the Flash Crash is not the only peril that awaits the unsuspecting investor nor is liquidity just a short term concern. Thus, in order to address the long term risks of investing in the financial markets, I shift the focus to macroeconomic crashes, specifically currency crises in developed markets. Decades of early warning system literature produced several empirical models to predict currency crises where most models focus solely on macroeconomic variables. However, it takes time for an economy to reflect the fragilities of the system in macroeconomic variables. Moreover, typical macroeconomic variables relate to the liability side risks of government balance sheets. The global financial crisis has shown us that asset side risks are just as important. In order to address this gap, I include liquidity related market variables as proxies for asset side problems. The significance of market variables in successfully determining currency crises suggests that liquidity is the primary determinant of crises both in the short and the long run. This major role liquidity is found to play in currency crashes is another novel contribution of this thesis to the existing literature.

All in all, despite the diverse nature of topics covered in this thesis, a twofold motivation governs the whole research. The first goal of this thesis is to regain normality for financial return distributions via subordination while the second is to predict financial crashes of varying time horizons.

1.2 Structure

Chapter 2 covers several theoretical and empirical market microstructure models to evaluate the influence of key components that derive asset prices (Easley and O'Hara (1992); Kyle (1985); Veronesi (1999)). The effects of market microstructure on normality of asset returns and realized variance is examined (Epps (1979); Zhang, Mykland and Aït-Sahalia (2005)) and optimum sampling strategies are reviewed (Bandi and Russell (2008); Aït-Sahalia et al. (2010)). Alternatives to normally distributed asset returns and the applicability of time changed Brownian motion is assessed.

The subordination approach introduced in Chapter 2 diverges from the literature on many fronts. For an extensive high frequency dataset, I start by rebuilding the order book for a selected number of stocks and use the information contained within the order book to recover the normality of asset returns via subordination under transaction time, a process I denote as "natural time".

In Chapter 3, I build upon the lessons learned from Chapter 2. Specifically, order book information, which was found to be influential in determining volatility, is used to predict episodes of sudden price dips in a high frequency setting. Consequently, Chapter 3 focuses on key order book components suggested by information-based market microstructure models to obtain a robust flash crash identification measure.

Furthermore, a novel nonlinear liquidity based crash prediction metric, "market heat", is proposed and tested against a linear and a volume-based crash predictor using linear discriminant analysis.

Finally in Chapter 4, crash prediction techniques are extended into early warning systems in order to predict large scale currency devaluations, which destabilize economies and depress growth for years. Existing early warning system literature primarily focuses on emerging markets as developed markets have long since been regarded as not susceptible to wild currency fluctuations. The global financial crisis of 2008 showed us otherwise. I aim to fill this gap in the literature by focusing on developed markets.

Both existing binary models such as the signaling approach of Kaminsky, Lizondo and Reinhart (1998) and the multivariate model of Berg and Patillo (1999), and panel estimations are employed in this chapter. In addition to an array of macroeconomic variables, market variables related to the global banking system are also included in all estimations. Using a crash threshold of 2% loss and a 1-month forecast horizon, several binary and panel models are estimated.

Chapter 2: Normally Distributed High Frequency Returns

2.1 Introduction

In this chapter, I aim to find the variables that cause the empirical deviation of financial returns from normal distribution and use these variables as subordinators to achieve normal returns. The findings presented in this chapter support the use of subordination as a method of achieving normality in addition to identifying several order book variables that can be used to account for heteroscedasticity. As such, the natural time approach introduced in this chapter achieves normality on several accounts and contributes to the literature by offering a new way to approach high frequency returns. Thus, Chapter 2 fulfills the first goal of this thesis, by attaining normally distributed returns. Moreover, Chapter 2 also provides a set of new variables which may be used to predict high frequency crashes, part of the second objective of this thesis. The ability of these variables to account for flash crashes is later put to test in Chapter 3.

The normal distribution assumption is central to many financial theories. However, empirical results, especially for high frequency data, often provide evidence against the normal distribution assumption (Müller et al. (1990), Dacorogna et al. (2001)). Excess skewness and kurtosis as well as price jumps cannot be justified within the normal distribution framework (Merton (1976); Taleb (2008)). Various different distributions have been suggested in its place but none can practically account for the peculiarities of financial returns. The additional microstructure effects observed in high frequency financial series added to these deviations from normality make the consolidation of these aspects under a unified framework even harder.

In this chapter, I provide an alternative explanation to the empirical divergence of financial returns from the normal distribution. The first key observation one needs to make when evaluating the distribution of asset returns is that most statistical analysis in this area is conducted using a physical time approach. However, the superimposition of a time grid on the transactions distorts the actual timing of trades. A second factor that is often overlooked is how the environment in which the prices are formed, specifically the order book imbalances, evolves over time.

The "natural time" approach that is detailed in Section 2.3, addresses these two key observations and aims to test the validity of normal distribution under a high frequency setting. Under the natural time approach, instead of sampling in physical time, transaction time[2] is used to record each trade as it materializes. By moving to the tick time sampling, the need to force each trade into a time slot is removed as one does not need to force the trades into predetermined sampling points as in calendar time. Additionally, when using calendar time sampling, methods of diurnalization is often employed to remove deterministic intraday patterns. Such deterministic patterns are usually observed during market open and close where number of trades and volume of trades spike. Sampling in calendar time cumulates the considerable trade information observed during these intervals into a handful of data points which then manifests itself in deterministic intraday patterns.

[2] Transaction time and tick time are used interchangeably throughout Chapter 2.

Sampling in transaction time; however, allows the sampling frequency to increase (decrease) as trades materialize faster (slower) producing variable number of data points. Thus, instead of removing the information contained in trades via diurnalization, transaction time retains this information by sampling according to the trading intensity which in return produces comparable data points.

Natural time approach also addresses the trade “environment”. By focusing on the factors through which prices are formed, the seemingly erratic behavior of volatility is accounted for. Variables derived from the limit order book are used to a form a gauge for volatility, which is used to subordinate raw returns, resulting in normally distributed return series. Hence, the goal of this chapter can be summarized as finding the best approximation for the “natural time” that results in normally distributed subordinated returns. The choice of sampling frequency and variables used in the subordinator function are the key to the success of this method.

Put simply, subordination based studies take variance related order book information to create an “instantaneous” volatility gauge, which is used to transform the original time series. Previous calendar time based subordination studies have found volume and number of trades to contain volatility related information such that normality could be recovered under certain periods (Clark (1973); Ané and Geman (2000); Silva and Yakvenko (2007); Velasco-Fuentes and Ng (2010)). Corresponding variables under the transaction time sampling, namely volume and duration, are used here as well. However, by using only these variables, the literature has neglected important information contained in the order book which can be used to explain the price formation process. For this reason, in addition to volume and duration, order book variables such as the imbalance in the standing order book and the difference in the number of bid and offer initiated trades are used to augment the models mentioned above. Asymmetric versions of the same subordination procedure are also tested.

The natural time approach is applied to 10 highly liquid LSE listed stocks for 4 quarters each. In several cases normality of returns is achieved. Since subordination essentially accounts for the volatility in the data, the natural time approach is tested against the standard GARCH(1,1) model. Natural time is found to dominate GARCH results with respect to normality, with the exception of a single instance. These findings suggest that volatility can be modeled efficiently under tick time sampling so much that subordination results in normally distributed returns.

The results in this chapter support the normal distribution assumption that is central to finance. However, they also point to the changes one needs to make in the standard model such as the sampling methodology. In addition to providing evidence for the normal distribution assumption, this research contributes to the literature by focusing on order book variables which contain relevant information that may be used to forecast volatility. The variables found to be influential here can be employed by market players to adjust their leverage or by financial regulators to assess the health of market. Either use will contribute to the efficiency of financial markets.

In the following sections, I will first take a closer look at how the financial markets operate and how various market microstructure effects contaminate the price evolution process (Mandelbrot (1963); Tauchen & Pitts (1983)). Key concepts such as information-based microstructure models and their implications on the use of duration between trades and trade size are examined. Stealth trading hypothesis and Kyle's λ as a measure of market resiliency are reviewed. The link between trade size and price impact is established. The effect of liquidity on absorption limits is evaluated. In addition to these conventional market microstructure models, seasonality and intra-daily patterns documented in the literature along with studies on the impact of scheduled macroeconomic announcements on risk premia are presented. The case for the use of a liquidity measure in accounting for market dynamics is strengthened by findings on post-announcement drift, overreaction and cascading effects.

Section 2.2.1 focuses key market microstructure models to identify the instrumental elements of the price process. The effects of homogenization and sampling techniques is studied via the vast realized variance literature and several calendar time and intrinsic time sampling techniques used are reviewed in Section 2.2.2. Drawing on the findings of realized variance literature, the need for dynamic sampling strategies, especially during high volatility states, becomes apparent. Hence, tick time is established as the sampling method.

To follow, Section 2.2.3 takes a closer look at time deformation and empirical studies that have employed subordination techniques to recover normality. Findings of the previous sections are then combined in Section 2.3 to create an alternative subordination approach, namely "natural time". Section 2.3.1 introduces stochastic time changes while Section 2.3.2 describes in detail the subordinators tested in Chapter 2. Section 2.3.3 gives details of the maximum likelihood estimation procedure while Section 2.3.4 introduces the evaluation methods assessing the distribution of returns. Section 2.4 introduces the dataset, shows the effects of sampling on returns and presents a supporting analysis of the variables used for subordination. The model results are presented in Section 2.5 and Section 2.6 concludes this chapter with a discussion of the results.

2.2 Literature Review

This three-part literature review aims to identify the two main components of the natural time approach introduced in this chapter, namely, the most appropriate sampling methodology for high frequency returns and a list of variables influential to the price formation process. Natural time approach draws upon the findings of this literature review to successfully achieve subordinated returns under tick time sampling.

The first subsection of this literature review focuses on how information is conveyed in financial markets. Several market microstructure models that explain trading patterns are reviewed and variables that effect price variance are identified. The natural time approach combines the variables presented in this subsection while accounting for variance related information. The second subsection reviews synchronization methods under physical time and identifies the inherent problems of working in the time domain. Alternate sampling methodologies are reviewed and the benefits of using tick time sampling, which forms an integral part of the natural time approach, are discussed. Finally, the last subsection reviews previous subordination based studies aimed at recovering normality. Natural time combines the variables identified in the first subsection under a stochastic subordination setting to recover normality of returns under tick time sampling.

2.2.1 Market Microstructure

The journey of quantitative finance starts with "*Théorie de la Spéculation*" where Louis Bachelier (1900) first applied normally distributed error terms to evaluate French stock options. This simple yet versatile stochastic process, Brownian motion, was later adapted to finance by Wiener (1923). The same idea of normally distributed price innovations was also used to create the infamous Black & Scholes (1973) option pricing formula. Lying at the heart of numerous financial studies, the assumption of normally distributed financial returns has been increasingly challenged. The assumptions of the efficient market hypothesis have been undermined by the microstructure manifestations observed in equidistantly time-spaced financial time series.

Several reasons emerge as responsible for the inability of the random walk model to account for empirically observed market dynamics. The lack of arbitrage, cash constraints, trading frictions and transaction costs, dependence of successive observations and non-stationarity are some of the key elements that contribute to the non-normality of empirical series. Transaction costs prevent arbitrageurs from instantaneously removing price discrepancies from financial markets, undermining the efficient market hypothesis. Cash constraints, on the other hand, may force market players to initiate stop-loss orders fueling price overshoots hence causing dependence of successive observations, fat-tails and non-stationarity, some of the key elements that contribute to the non-normality of empirical financial time series (Mandelbrot (1963); Fama (1965); Engle (1982); Bollerslev 1986)). Many of these market microstructure effects that underlie return anomalies have been documented in detail in the extensive microstructure literature (Aït-Sahalia et al.

(2010); Aït-Sahalia & Yu (2009); Admati & Pfleiderer (1988); Bandi & Russell (2008); Dacorogna et al. (1993); Glosten & Milgrom (1985)). Much focus has been given to bid-ask spread with two main strands of models, namely inventory-based and information based models. Inventory-based models argue market makers adjust their quotes to mirror their inventory positions, while information-based models focus on the costs associated with adverse selection.

2.2.1.1 Conventional Market Microstructure Models

Inventory-based models argue that market makers will adjust their quotes so as to mirror their inventory positions. As compensation for holding excess inventory in the face of adverse market movements and providing liquidity, the market makers demand the bid-ask spread (Bagehot (1971); Stoll (1978)). Alternatively, Roll (1984) has focused on order handling costs, calculating the effective bid-ask spread. He used the first order serial covariance to compute the average absolute value of price change when no new information has arrived in the market. Roll's specification shows that in times of higher uncertainty and hence wider spread, the effective trading costs increase for market participants while the market maker's profits swell as compensation for higher risk. Roll computed the effective spread as:

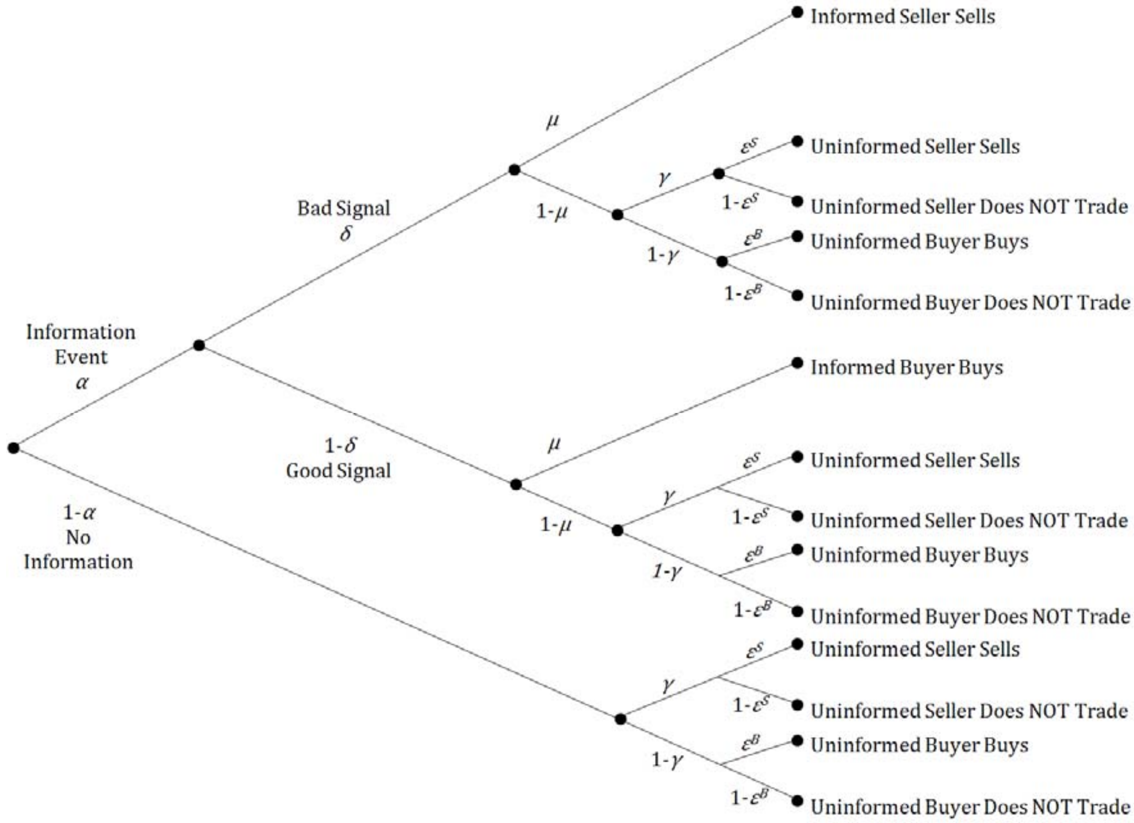
$$\text{Spread} = 2\sqrt{-cov}, \tag{2.1}$$

where cov is the first order negative autocorrelation.

Information-based models on the other hand focus on the costs associated with adverse selection. Glosten and Milgrom (1985) mapped the bid-ask spread as the market maker's tool against traders with insider information. In their heterogeneous expectations model, a signal Ψ with information on the value of an asset arrives at each time node. A negative signal arrives with probability δ and a positive one with $(1-\delta)$. The asset assumes a low value of V^{-}given a bad signal and a high of V^{+} otherwise. Two types of market agents constitute the market, namely uninformed traders with only public information and informed traders with knowledge of the true asset value. A priori percentage of insider traders present in the market is denoted by μ and the probability that an uninformed trader buys or sells is given by γ_B and γ_S, respectively. Finally, every market agent completes a unit transaction at each time node. Given this setup, the market maker revises its quotes via Bayesian updating gradually revealing insider information given its order flow.

Easley and O'Hara made two important extensions to the original Glosten and Milgrom model. In their Easley & O'Hara (1987) model, they have introduced the possibility of no information with probability α. Additionally, the uninformed traders were allowed to trade small and big quantities where X_B^1, X_B^2, X_S^1, X_S^2 denote respective probabilities. The second model introduced in 1992 removed differences in trade quantities while allowing uninformed traders not to trade with a probability of $(1-\varepsilon)$. For both models, information signal Ψ arrived once before the trading day. The figure below outlines the setup of Easley and O'Hara (1992) model:

Figure 2.1: Easley & O'Hara – 1992 Model



Despite their shortcomings, such as constant percentage of informed traders determined a priori, the asymmetric information models of Easley and O'Hara underscore several important market dynamics, where trade size, duration between consecutive trades and lack of trades reveal information about the latent price dynamics. Further evidence on effects of trade size on price evolution can also be found within the stealth trading hypothesis. The impact of trade durations of return series will again be examined while looking at alternative procedures for time deformation. Additionally, the Easley and O'Hara (1992) model will be used in flash crash identification in conjunction with the subordination variables identified in Section 2.2.

Information-based models also helped pave the way to disentangle permanent and transitory components of transactions on price processes (Biais, Glosten and Spatt (2005)). Kyle (1985), for example, mapped the evolution of asset prices and insider trading quantity as an equilibrium model. For convenience reasons, notation from Instefjord (2005) will be used below. In his model, Kyle assumes asset returns and the quantity noise traders transact to be normally distributed such that:

$$x \sim N(0, \Sigma^2), \tag{2.2}$$

$$y \sim N(0, \sigma^2), \tag{2.3}$$

where x and y represent the asset price and the quantity traded by noise traders.

One shortcoming of Kyle's model is the violation of non-negativity constraint for asset prices, which is infeasible for stocks due to their limited liability nature. However, as the insiders act on their private information on the true value of the asset; it is not the asset price but the difference between the market clearing price and the latent price that determines their profits. Hence, negative asset prices do not undermine the validity of Kyle's model given this profit-based perspective.

Both noise and informed traders submit only market orders to a single auctioneer which observes an aggregate quantity $q = y + z$, where z is the quantity demanded by informed traders. The auctioneer then sets a clearing price p with a zero return expectation. The clearing price the auctioneer sets is given by:

$$p = E[x \,|q]. \tag{2.4}$$

An equilibrium exists such that:

$$z = \beta x, \tag{2.5}$$

$$p = \lambda q, \tag{2.6}$$

where β and λ are constants.

Given the insider trader's profit function:

$$E[\pi(x)] = z(x - dz), \tag{2.7}$$

it can be shown that

$$z = \frac{\sigma}{\Sigma} x. \tag{2.8}$$

Equation (2.8) suggests that aggressiveness of insider traders is correlated with the ratio of noise trading dispersion and asset price standard deviation. These models also gave rise to the "stealth trading hypothesis", where market participants with insider information try to avoid information leakage while submitting orders. Insiders are forced to find a balance between the risk of effecting prices adversely with block trades - impact risk - and price risk due to order slicing.

The impact of order size has been studied in a linear setting by Bertsimas and Lo (1998), Almgren and Chriss (2000). Barclay and Warner (1993), Chakravarty (2001), Cai, Ouyang and Wong (2011) and Huang (2011) found evidence of stealth trading in stock and option markets where medium sized trades tend to move the prices the most. Moreover, Anand et al. (2005) examined the evolution of liquidity and find institutional medium sized orders to be informed. The authors also find a behavioral difference in the actions of institutional traders where they use aggressive market orders to exploit their informational advantage, absorbing liquidity in the morning and acting as liquidity providers with unaggressive limit orders in the afternoon. Similarly, Blau (2009) suggests that stealth traders adjust their order size with respect to the market depth. Malik and Ng (2009) also find evidence in support of the information-based microstructure theory, where the bid-ask spreads for FTSE100 stocks tighten during the day. Informed trader aggression often exhibits itself in the volume of trades. As such volume will later be used as an essential component in the time deformation process as the stealth trading hypothesis shows that the volume of trades affect the price formation process.

Kyle (1985) also identifies three major components to liquidity, namely tightness, depth and resiliency. Given this setup "Kyle's λ" becomes a measure of market sensitivity to transaction size, where orderbook imbalances can be used to infer impact of order size (Aldridge (2010)). The price impact of orders can be represented as:

$$\Delta P_t = \alpha + \lambda\, OBI_{\mathrm{t}} + \varepsilon_{\mathrm{t}}, \tag{2.9}$$

where OBI_{t} is the orderbook imbalance computed as the difference between the bid and ask quotations.

Extensions to Kyle's λ have been suggested by Amihud and Mendelson (2000), who find that illiquidity is priced into return expectations. Large (2007) mapped resiliency of the limit orderbook for Barclays shares using a continuous multivariate point process and found that in less than 40% of the cases the orderbook could replenish itself within a half life of 20 seconds. Ng (2008) tested the absorption limits of financial markets within a nonlinear ACD framework and reported that markets are incapable of absorbing large block trades introducing additional "time costs of liquidity". These findings regarding the resiliency of financial markets strongly support the stealth trading hypothesis, where market participants actively try to balance liquidity and information costs, and necessitate the need to use some form of liquidity measure in order to account for high frequency dynamics.

2.2.1.2 Non-Conventional Market Microstructure Models

Additional microstructure effects have surfaced with greater availability of high frequency data, further revealing the seasonality in returns. Yearly, monthly and weekly deterministic patterns have been documented by French (1980), Gibbons and Hess (1981), Apolinario et al. (2006) among others. Similarly, Engle and Russell (1998) developed the autoregressive conditional duration (ACD) model to account for deterministic diurnal trading patterns such as consistent high volatility observed at market open and close.

The effects of scheduled macroeconomic announcements on diurnal return and volatility was another key area of research that flourished. The literature on effects of scheduled announcements has been deeply influenced by the canonical work of Veronesi (1999).

In his rational expectations model, Veronesi allows investors to hold a risk-free or a risky asset whose dividend returns are given by:

$$dD = \theta_t d_t + \sigma \, d\omega, \tag{2.10}$$

where θ_t and $d\omega$ denote the state variable and a Wiener process respectively.

The state variable follows a two-state continuous-time Markov regime-switching process and can assume values of $\overline{\theta}$ and $\underline{\theta}$ with a transition probability matrix between time t and $t + \Delta$:

$$P(\Delta) = \begin{pmatrix} 1-\lambda\Delta & \lambda\Delta \\ \mu\Delta & 1-\mu\Delta \end{pmatrix}, \qquad (2.11)$$

where $\overline{\theta} > \underline{\theta}$.

Given this setup Veronesi showed that investors' overreaction to bad news in good times and underreaction to good news in bad times stems from state uncertainty, where investors demand a premium for bearing additional risk. Savor and Wilson (2015) detect almost an annualized 10% excess returns for announcing stocks compared to non-announcing ones. Their results confirm the state dependence reasoning for increased risk premia. Savaşer (2011) finds evidence in support of Veronesi's hypothesis with price contingent stop-loss and take-profit orders surrounding scheduled announcements. She also underscores the effects of the orderbook imbalances, which account for a substantial portion of the news announcement effects. Despite their orthogonality to news, series of stop-loss/take-profit orders may create a positive-feedback mechanism that moves prices in a given direction. This cascading effect has also been documented by Osler (2005). These findings highlight the role of order book imbalances in accounting for news effects.

Similarly, Andersen et al. (2003) find that the mere presence of scheduled announcements increases volatility independent of the news surprise component. Using a 5-minute sampling time, Andersen et al. (2003), Andersen et al. (2007) and Harada and Watanabe (2009) document an almost instantaneous price adjustment to news producing "jumps" while volatility adjusts gradually to the new information. Studies with a higher rate of sampling however, produce dissimilar results. Using 1 minute prices of German Bund futures, Hautsch et al. (2011) show that post announcement drifts continue for minutes after the news release. The authors dissect volatility into noise and efficient components, both of which is found to be significantly affected by “net order flow”. The noise component of volatility reaches a peak 10 minutes before the announcement due to drying liquidity and jumps further following big surprise announcements. The reversal of noise volatility to pre-announcement levels 10 minutes after such releases is also suggestive of overshooting effects.

Overshooting and counter reactions have also been documented by Entorf et al. (2009) on a different sampling scale. Using 15 second Xetra DAX returns, the authors identify counter reaction patterns to ifo[3] and ZEW[4] releases which manifest themselves after 30 and 45 seconds following the announcements. Glattfelder, Dupuis and Olsen (2011) on the other hand, employ an intrinsic time approach to map the overshooting behavior in FX markets and develop several scaling laws.

---

[3] ifo Business Climate Index, reported monthly by the Ifo Institute of Economic Research, is a seasonally-adjusted leading indicator of German business activity. The index is constructed based on approximately 7,000 surveys distributed among businesses in manufacturing, construction, wholesale and retail sectors. Businesses are asked to qualitatively assess the current business conditions (good/satisfactory/poor) and provide their expectations for the next 6 months (more favorable/unchanged/less favorable). The surveys are weighted according to industry importance and the balance value is calculated by taking the percentage difference between the positive and negative responses. The index is formed by the seasonally-adjusted geometric mean of balances normed to the base year. For further details please refer to http://www.cesifo-group.de.

[4] ZEW Indicator of Economic Sentiment is a monthly economic survey that reflects the expectations of up to 350 financial analysts. Contributing experts are asked to evaluate the state of the German economy within the next 6 months on a qualitative scale (optimistic/no change/pessimistic). ZEW is then computed as the percentage difference between optimistic and pessimistic responses. For further details please refer to http://www.zew.de.

The existence of scaling laws hints at systematic microstructure effects and provides insights into duration and number of transactions' effects on the price processes. On the whole, studies on scheduled macroeconomic announcements suggest that the incorporation of order book imbalances and market liquidity is of paramount importance in understanding the volume and price tides observed before and after news releases.

The theoretical and empirical market microstructure models presented in this section underscore the importance of several parameters which are essential in accounting for market movements. The scheduled macroeconomic announcement studies show the drying of liquidity and a sudden spike right before and after new releases respectively in addition to occasional price jumps.

Moreover, the stealth trading hypothesis shows that volume of trades determine their market impact. Kyle's measure of resiliency and the time it takes a market to recovery from a large block trade shows that several frequent block trades could put a market out of balance. Thus, not only the volume of trades but also the market's absorption limit or in other words the prevailing liquidity conditions affect the evolution of prices.

Announcement reactions, scheduled or unscheduled, bring another dimension to the price process. The intensity of trades following sudden changes of sentiment reflects an inevitable herding behavior following important news. The number of transactions spike and important adjustments to asset prices are realized during these short time intervals. This inherent correlation between number of transactions and return variance has been previously tested in a physical time setting (Ané and Geman (2000)). I will follow a similar approach here that will allow me to move to an alternate time frame. Instead of cumulating the number of transactions during a given time interval, I will cumulate the number of time units between a given number of transaction. Then, for a given time interval, a high (low) number of transaction will be the equivalent of short (long) trade durations.

Hence, three important components affecting price evolution emerge from this section, namely volume, market liquidity-imbalance and duration between trades. The use of these key components under a unified framework will be the key contribution of this work forming a comprehensive approach accounting for most if not all market dynamics. The specific format in which these market variables will be used to recover normality of asset returns will be clearer in the following sections.

2.2.2 Realized Volatility & Optimal Sampling

In this subsection, the price evolution of financial assets will be mapped within a Brownian motion framework. The market microstructure effects that were outlined Section 2.2.1 are introduced into the observed financial time series data and the effects of market microstructure on optimum sampling frequency from a realized variance standpoint are examined. Several calendar time sampling and homogenization techniques along with methods for removing deterministic seasonality in financial series equally spaced in calendar time are presented. Finally, the use of tick time and its ability to account for market speed and seasonality is considered.

The statistical theory suggests that sum of the squared errors sampled at increasingly high frequencies should in probability converge to the realized variance (RV) of the latent quadratic variation. In a continuous stochastic setting let S_t denote the efficient price of a security which follows a geometric Brownian motion as below:

$$dS_t = \mu S_t d_t + \sigma S_t dW_t, \tag{2.12}$$

where S_t represents the asset price at time t, μ is the drift component which is often set to 0 since drift is negligible at high frequencies, σ is the volatility of diffusion process, a strictly positive càdlàg process, and dW_t is a Wiener process.

Alternatively the price evolution process for the log-price can be summarized as an arithmetic Brownian motion:

$$X_t = \mu d_t + \sigma dW_t, \tag{2.13}$$

The integrated variance of the latent price process can then be approximated by:

$$[X, X]_T = \sum_{t_i} \left(X_{t_i} - X_{t_{i-1}} \right)^2, \tag{2.14}$$

since

$$[X, X]_T \xrightarrow{p} \int_0^T \sigma_t^2 dt, \tag{2.15}$$

as the sampling interval d_t approaches 0 (Zhang, Mykland and Aït-Sahalia (2005)).

However, sampling at higher frequencies comes at a cost. In reality the price process one observes in the market is heavily contaminated by various types of market microstructure effects. Thus, the realized variance calculated using high frequency data diverges from its true value. Epps (1979) first documented the substantial decrease in cross-correlations between stocks at increasing sampling frequencies. His findings were later complemented by Lundin, Dacorogna and Müller (1998) and Tóth and Kertész (2009). Factors that contribute to the Epps effect include bid-ask spread, price discreteness, jumps, asynchronous trading, infrequent trading, decimalization, informed trading among others. Münnix, Schäfer and Guhr (2010) for example find that discretization can account up to 40% of Epps effect, especially for lower valued stocks.

Now let us assume that the observed price process, X_{t_i}, is the sum of the latent efficient price process, Y_{t_i}, plus an error term, ε_{t_i}, which incorporates all microstructure based effects. The observed price process is then:

$$X_{t_i} = Y_{t_i} + \varepsilon_{t_i}, \tag{2.16}$$

where ε_{t_i} is an i.i.d. white noise process.

Given the above setup, the realized variance of the observed process then becomes:

$$[X, X]_T = [Y, Y]_T + [\varepsilon, \varepsilon]_T, \tag{2.17}$$

since the cross product term, $2[Y, \varepsilon]_T$, cancels out due to independent noise assumption. Several studies in the realized variance literature relax the i.i.d. assumption as well.

The reason why the sum of the squared returns for the observed price process is an inconsistent estimator of true volatility becomes clear in Equation (2.17). The orders of magnitude for the two components differ with $[Y, Y]_T = O_p(\sqrt{d_t})$ and $[\varepsilon, \varepsilon]_T = O_p(1)$. In simpler terms, the variance of the error term dominates the variance of the latent price process at high frequencies.

Realized variance literature exclusively focused on this behavior of financial series in order to find an optimum sampling frequency that balances the adverse effects of microstructure noise with the gains of frequent sampling. Various parametric and nonparametric approaches have been employed in the literature along with different time sampling schemes. Zhang, Mykland and Aït-Sahalia (2005) modeled the first nonparametric consistent estimator of realized volatility. They combined a sparsely sampled RV estimator with one that uses all available data to come up with an efficient two-scale estimator. Zhang (2006) expanded their findings into the multi-scale dimension. While Barndorff et al. (2011a) and Huang and Lee (2013) used subsampling to overcome microstructure effects, Barndorff-Nielsen et al. (2011b) employed a kernel-based parametric approach to attain the same convergence rate as the multi-scale estimator. Aït-Sahalia, Mykland and Zhang (2005) also showed that their parametric estimator is robust to Gaussian error misspecification.

Similarly, Bandi and Russell (2008) used calendar time and mid-quotes to evaluate the utility of optimal sampling in their bias correction framework. They find that the optimum sampling interval in physical time varies within their dataset and the ad hoc 5 minute sampling employed often in the literature, Andersen et al. (2001), actually conforms with the their optimum sampling interval. In their study they also advocate the use of mid-quotes as they would be less prone to bid-ask bounce effects. However, Hansen and Lunde (2006) suggest both in calendar time and tick time the mid-quotes are subject to further contamination due to non-synchronous updating of the bid and ask prices when prices move in a given direction. They also document noise-efficient price dependence in both time scales and find that its takes approximately 10 ticks for dependence effects to subside. Parallel to the findings of Bandi and Russell (2008), Oomen (2006) finds that the optimum sampling interval both for calendar based and tick based sampling to be dynamic in a pure jump setting. Oomen also underscores the fact that compared to calendar time sampling (CTS), transaction time sampling (TTS) is a better estimator of quadratic variation in the absence of noise. The results do not vary greatly with the introduction of noise as the loss function of CTS is heightened for high levels of intensity as well as increased volatility in the arrival intensity.

Let us now take a closer look at the sampling schemes employed in the above studies. Whether it is the calculation of covariance among different stocks or computation of realized variance for a single asset, synchronization requires data to be fit into some form of a grid. Especially, for high frequency time series, which are almost always unevenly spaced in physical time, synchronization is essential for statistical inference.

The RV literature has exclusively focused on such homogenization techniques due to their immediate effect on the optimum sampling frequency. Two major synchronization methods emerge in the literature for homogenizing high frequency series of a single asset in calendar time.

The "previous tick" method (Wasserfallen and Zimmerman (1985)) is perhaps the most frequently used method of transforming inhomogeneous tick data into evenly time-spaced homogenous data (Pagel, Jongh & Venter (2007); Zhang (2011)). One major shortcoming of this method; however, is spurious jumps observed in case of extended periods of missing data (Dacarogna, Gençay, Müller, Olsen and Pictet (2001)). The previous tick method can be summarized as below:

Let t_i be the successive homogeneous sampling intervals such that:

$$t_i = t_0 + i\Delta_t. \tag{2.18}$$

Then the associated prices are according to previous tick method are:

$$Z_i = Z_{t_i} = Z_{j'}, \tag{2.19}$$

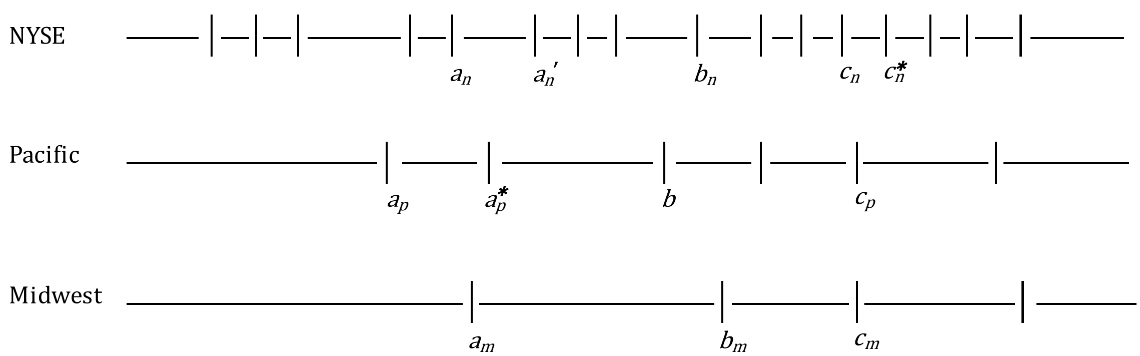
where subscript j and j' represent original and adjusted inhomogeneous time series.

As an alternative, "linear interpolation" forms the homogenous time series by interpolating between the nearest tick data observed just before and after the grid time (Müller et al. (1990); Andersen and Bollerslev (1997); Velasco-Fuentes and Ng (2010)). Although the difference between the two methods might be negligible, linear interpolation violates causality. As pointed out by Hansen and Lunde (2006) in a realized variance setting this interpolation scheme is not suitable since the quadratic variation of a straight line is zero in the limit. The linear interpolation scheme is outlined below:

$$Z_i = Z_{t_i} = Z_{j'} + \frac{t_0 + i\Delta t - t_{j'}}{t_{j+1'} - t_{j'}} \left(Z_{j+1'} - Z_{j'}\right). \tag{2.20}$$

Alternative approaches to the above major models which take into account multiple assets exhibiting asynchronous transaction data do also exist. Looking at co-integration in IBM stocks listed in different exchanges, deB. Harris et al. (1995) uses "replace all" - sometimes referred to as "refresh time" as in Barndorff-Nielsen et al. (2011b) - and "minspan" schemes. The procedures for both methods are similar. A price vector is formed by looking at successive time windows where each asset has traded at least once and adds the nearest previous transaction price for more frequently traded instruments. Figure 2.2 illustrates a stock traded in three difference stock exchanges asynchronously.

Figure 2.2: [5]



Tuple	Replace All	Minspan
1	(a_p, a_n, a_m)	(a_n, a_m, a_p^*)
2	(b_p, b_n, b_m)	(b_p, b_n, b_m)
3	(c_n, c_p, c_m)	(c_p, c_m, c_n^*)

The replace all method forms a new tuple once the stock has traded in either one of the three stocks and adds one trade each from the other stocks as soon as they are formed. Hence, the replace all method cannot adjust tuples by using information from the future. On the other hand, the minspan method creates its tuples by minimizing the time between trades included in the vector by allowing for trades that occur after the limiting trade. Hence, while the replace all method would form its first vector by sampling the trades (a_p, a_n, a_m), minspan replaces the trade that occurred at time a_p with a_p^* forming the vector (a_n, a_m, a_p^*). A shortcoming of both sampling schemes; however, is their dependence on the frequency of the least traded asset which results in throwing away of a major portion of the available data.

[5] Figure adopted from deB. Harris et al. (1995), page 6.

Additionally, Aït-Sahalia et al. (2010) proposed the "generalized sampling time", where an arbitrary tick data point is selected for each asset within a given time interval. The authors advocate such a procedure would also be robust to data misplacement errors given that misplacement occurs within each time interval.

As mentioned earlier, although the literature predominantly focuses on sampling in physical time, this is not the only option. Dacorogna, Müller, Nagler, Olsen and Pictet (1993) proposed the use of θ-scale, which accounts for intraday and intraweek deterministic patterns. Essentially the θ-scale removes seasonality of volatility due to the operating hours of the 3 main trading regions via a subordination process, Dacorogna, Gauvreau, Müller, Olsen and Pictet (1996). The θ-scale can be summarized as follows:

$$\theta(t) = a_0(t - t_0) + \sum_{k=1}^{3} \int_{t_0}^{t} a_k\,(t')dt', \tag{2.21}$$

where a_0 is the minimum market activity and a_k represents the effects of the three main markets to market activity, namely Europe, USA and Asia.

Akin to intrinsic time, the business time sampling scheme used in Oomen (2006) removes both deterministic and stochastic components of volatility in a pure jump setting by sampling based on the expected number of trades. Oomen applied the idea of constant jump intensity in business time to transaction time sampling which he finds to be superior to business time. Generally referred to as tick time, transaction time, accounts for trades as they materialize. Contrary to calendar time where trades are irregularly spaced, in tick time each transaction falls nicely on the tick grid. This property of tick time is quite advantageous as it inherently removes the effects of asynchronicity.

Further advantages of using tick time come afore given the intraday and intraweek behavior of volume and volatility of asset returns. Sampling in physical time often requires data to be adjusted for deterministic market patterns. Mostly in realized volatility or duration studies conducted, market volatility or volume are de-seasonalized with the "diurnalization" process where data is adjusted for the deterministic market patterns via the utilization of splines, Fourier transforms or kernel based estimators. Fourier transforms employed in Andersen et al. (2003) are very smooth process, which may not be able account for "jump" effects observed around scheduled macroeconomic announcement, whereas spline methods such as cubic spline employed in Engle and Russell (1998) are much more flexible. However, the choice of nodes for spline may yet present problems. Ng (2008) addresses the choice of kernel bandwidth with cross validation. Empirically diurnalization may produce satisfactory results but its exact effects on the object of interest is little explored (Martens et al. (2002); Allen et al. (2009)).

Despite the fact that the literature predominantly focuses on sampling in physical time, this is not the only option. Synchronization and diurnalization essentially aim to produce data points which are comparable. Sampling in tick time inherently eliminates the need for synchronization since the data points determine the sampling grid itself. Furthermore, by sampling at a fixed number of ticks, one completely avoids the processes of diurnalization as the clock moves faster (slower) when market activity is high (low).

Given its advantages in adjusting for market seasonality (Oomen (2006); Dacorogna et al. (1993)), tick time will be used in this chapter while evaluating the distribution of stock prices in the high frequency setting. The importance of tick time will be much more apparent in the next section while normalizing financial time series via subordination.

Unlike other sampling schemes presented in this section, tick time will not require diurnalization, hence will not introduce any additional calendar time related errors into the time series.

2.2.3 Alternate Distributions & Time Deformation

This section will briefly look at the alternative distributions suggested for financial time series which often diverge from normality. Most importantly, subordinated Brownian motions will be considered and their application to the financial returns will be examined. Then main drivers of microstructure effects detailed in Section 2.2.1 and tick time sampling methodology explained in Section 2.2.2 will be joint under a subordination structure that will be used recover normality of asset returns. Finally, an empirical application of the subordination scheme will be outlined for spillover effects observed in the stock market.

The empirical divergence of asset returns from normality, excess skewness and fat tails, has long spurred interest in alternate distributions such as the exponential and *t*-distribution. Merton (1976) proposed the addition of jumps to the original continuous stochastic diffusion process in Black and Scholes (1973) to account for fat tailed asset returns. Tauchen and Pitts (1983) explored the possibility of normal mixture distribution, while Mandelbrot (1963) examined the stable distribution. Mandelbrot posited that although asset returns were approximately independent they were characterized by unbounded second moments and advocated the use of stable Paretian distribution. One empirical shortcoming of the stable Paretian distribution from a practitioner's perspective; however, is the lack of closed form distributions which necessitates numerical estimation via the characteristic function. Mandelbrot and Taylor (1967) later suggested that the stock price distribution could also be normal under subordination, where the subordinator has an infinite mean and variance stable distribution. However, substantial evidence against unbounded first and second moments undermines the applicability of stable processes to financial return series. Perry (1983) and Cont (2001) find that variance of returns do converge to a finite value for US and French stocks. Perry further concludes that it might be the "complex fashion" volatility evolves that causes the fat-tailed distribution we observe in financial return series, the usual suspects being nonlinearity, time and state dependence.

The notion of subordination can be captured by first looking at the no arbitrage assumption and Girsanov's change of measure often used in the derivation of Black-Scholes option pricing formula. In a no-arbitrage setting, the discounted asset prices form a martingale under the risk neutral Q-measure. It follows directly from this result that asset prices are semimartingales under the equivalent P-measure. Given Monroe's (1978) extension of Dubins-Schwartz theorem, any semimartingale can then be expressed as a "time-changed" Brownian motion. For a study on the evolution of time changes and subordination see Geman (2005).

Clark (1973) was the first to apply the subordination process to assets prices to recover normality of asset returns. He conjectured that financial return series, which are semimartingales, could be defined as subordinated Brownian motions such that:

$$X_t = W(\tau_t) \sim N(\mu, \sigma), \tag{2.22}$$

where process τ_t is the directing process or subordinator.

In Equation (2.22), the process $W(\tau_t)$ is subordinated to the original price process X_t and the subordinator τ_t is a càdlàg process that measures market's intrinsic time which apparently flows at variable rates. Clark (1973) tested the applicability of trade volume as a subordinator for cotton futures and found evidence in favor of the Gaussian distributed asset returns within an i.i.d. subordinator increments setting using cumulative trade volume. Karpoff (1987) also documented the connection between large trades and large price swings and conjectured that it might be linked to both factors' shared link to the underlying information process. Do et al. (2014) also found evidence of a strong link between trading volume and heteroscedasticity in asset returns.

Ané and Geman (2000) generalized the subordination framework by relaxing Clark's i.i.d. assumption in a finite variance jump setting. Using 1,5,10 and 15 minute sampling frequencies, Ané and Geman (2000), find transaction frequency to be a better subordinator compared to volume for S&P future contracts. Geman (2002) has also shown that the directing process can also be interpreted as the "mixing factor" within a normal mixture distribution setting, an often used distribution to account for excess skewness and kurtosis in stock returns. Murphy and Izzeldin (2006); however, questioned the reliability of moment estimation methods in Ané and Geman (2000) and presented counter evidence on recovery of normality using re-centered number of trades or volume. Silva and Yakovenko (2007) also used the number of trades as a subordinator using intraday tick data for Intel stock. Silva and Yakovenko (2007) find that an approximately Gaussian return distribution can be obtained using sampling frequencies that range from over 30 minutes to almost 3 hours. However, sampling at such sparse intervals makes the contribution of subordination over aggregation questionable.

Similar to Ané and Geman (2000), Huth and Abergel (2012) used the number of transactions to subordinate the returns for multiple assets. In a multivariate framework, Huth and Abergel (2012) chose to sample each time a trade occurs in any one of the assets creating a "common stochastic clock". Then by subordinating with an event time N, which represents the total number of trades in all assets under consideration, they obtained results that support normally distributed returns for 4 asset pairs. However, the large number of trades Huth and Abergel (2012) have used to obtain normality, which in one case reached almost 6,000, and the fact that the joint stochastic clock used only produces reliable results if the asset pairs have similar trading patterns suggest that their findings may be mostly attributed to aggregation.

Velasco-Fuentes and Ng (2010) further investigated the use of volume and number of trades as stochastic time changers. In a study using FTSE-100 futures tick data they have tested cumulative volume, total number of trades and their linear and quadratic combinations to recover normality. They have also explored the possibility of asymmetric market response to the sign of returns in order to reduce skewness. Using first and second order functions of volume and number of trades Velasco-Fuentes and Ng recover normality in two of the four sub-periods.

First part of my thesis will be closely related to the works of Clark (1973); Ané and Geman (2000) and Velasco-Fuentes and Ng (2010), aiming to recover normality of asset returns via the use of stochastic subordination. The contribution of this research is twofold. First, it extends the arsenal of possible factors that are most closely related with information arrival and intrinsic time. Second, tick time is used for the first time in subordination literature to test the assumption of normally distribution of financial asset returns.

2.3 Methodology

In this chapter, I take an atypical approach to stochastic subordination. Unlike its predecessors in subordination literature, which sample data in calendar time, this research is conducted under tick time. Hence, the applicability of the normal distribution assumption is tested for the first time under transaction time sampling, a major contribution of this chapter. When sampling in tick time, daily deterministic patterns present under physical time need not to be removed via diurnalization. Moreover, additional errors introduced while conforming to a calendar time grid is no longer present under tick time as comparable data points will fall onto the tick grid perfectly.

Furthermore, by using transaction prices and their returns, I also avoid using quotes which may react asymmetrically during unidirectional market swings. Asymmetric quote updates occur when market participants fail to update their bid (ask) orders when there is a rapid price increase (decrease). Such rapid price changes leave little time to market players to revise their orders. In return, market players tend to update the orders on the most “urgent” side of the order book. Thus, during a rapid price decrease, a market player would be more concerned to update his bid orders rather than any ask order. This phenomenon then manifests itself in the data as non-synchronous updating of quotes.

When compared with bid, ask or mid quotes, actual transaction prices, which are prices both the buyers and the sellers have already agreed upon, are better indicators of financial value of assets. Hence, log-returns calculated from transaction prices are not subject to microstructure contaminations such as non-synchronous updating of quotes.

Four important components affecting price evolution emerged from previous sections, namely volume, duration, market liquidity and order imbalance. The ability of these variables to successfully subordinate high frequency returns under tick time to achieve normality will be put to test. I will account for various market dynamics by extending the arsenal of possible factors that are most closely related with information arrival and intrinsic time.

Volume, as per its impact to push prices in a given direction is the first of these factors. such as Clark (1973) and Ané and Geman (2000) have found support for volume as a subordinator while sampling in calendar time. Similarly, Huth and Abergel (2012) and Velasco-Fuentes and Ng (2010) used the number of transaction to subordinate returns. However, as shown in Gillemot, Farmer and Lillo (2006), volume and number of trades cannot totally account for the volatility observed in the stock markets. This may be caused by the imperfect correlation these variables have with the latent process which drives volume, number of trades and volatility. Hence, as per the findings of information-based models, duration between trades is also added to the subordination framework to account for the speed with which market participants act in physical time.

The use of duration is new to the subordination literature and augments the model in two respects. Given the stealth trading reasoning presented in the previous sections, and the information-based market microstructure models, the duration between trades not only helps capture the speed of the market in real-time, but also reveals the private information content. By including duration between trades I allow physical time related information to be included while sampling in tick time.

In addition to the explanatory variables $volume$ and $duration$, proxies for the liquidity component of the market are included in my model, namely, net traded volume imbalance and net initiator imbalance. Net traded volume imbalance is the volume difference between bid and ask initiated trades – denoted as $Volume\ Imbalance$ or $Vol\ Imb$ – can be expressed as:

$$VolImb = Vol_{Bid} - Vol_{Ask} \ , \tag{2.23}$$

where Vol_{Bid} and Vol_{Ask} is the volume of bid and ask initiated trades, respectively.

Net initiator imbalance, on the other hand is the difference between the number of aggressors on buy and sell sides – denoted as $Initiator\ Imbalance$ or $Init\ Imb$ can be defined as:

$$InitImb = Num_{Bid} - Num_{Ask} \ , \tag{2.24}$$

where Num_{Bid} and Num_{Ask} is the number of unique bid and ask initiated trades, respectively.

Huang (2011) previously found evidence for a contemporaneous relationship between order imbalances and asset returns while looking at stealth trading in NASDAQ stocks. However, the bulk classification of trades using Lee and Ready (1991) algorithm and consistent buying pressure within their dataset renders Huang (2011)'s findings open to question. In this chapter, the explanatory power of order book imbalances will be put to test using high frequency trades that are perfectly classified into buyer or seller initiated trades via corresponding stock exchange codes. Furthermore, the effect of imbalances in the limit order book is also tested via the $Imbalance$ variable (the difference between standing bid and ask orders).

The addition of liquidity variables sets the scene in which the trades occur and adjusts for the impact of block or frequent trades given market depth or resiliency. However, it is highly unlikely for liquidity conditions to affect prices like volume of trades, where one consistently drives prices while the other acts as a determinant (multiplier) of price impact for a given trade. For this reason, the effects of order book imbalance on the price process is likely to be nonlinear. I will test this assumption during the subordination process.

By including these possibly omitted variables in the subordinator, I aim to regain normality of asset returns during all states of the world, without any need of additional adjustment to the data such as diurnalization. The use of an asymmetric response function similar to the one in Velasco-Fuentes and Ng (2010) is also examined. Thus, in addition to linear combinations of the three factors identified, the importance of nonlinear models will also be tested, given the inability of linear models in explaining asset price fluctuations.

2.3.1 Stochastic Time Change

The stochastic time change that will be applied to the raw return series can be described as follows. Define the price series of an asset sampled in calendar time as:

$$P_{cal}(c) = \left(P(c_1), P(c_2), P(c_2), \dots, P(c_{n-1}), P(c_n)\right), \qquad (2.25)$$

where c_i present sampling in calendar time.

Similarly define a stochastic (parent) process:

$$W(q) = \big(W(q_1), W(q_2), W(q_3), \cdots, W(q_{m-1}), W(q_m)\big), \tag{2.26}$$

where q denotes market's intrinsic time, the variable rate at which market activity flows. The stochastic parent process, W, is Brownian Motion in my case.

If a strictly increasing stochastic process:

$$s(c) = \big(s(c_1), s(c_2), s(c_3), , \dots, s(c_{n-1}), s(c_n)\big), \tag{2.27}$$

where $s(c_{i+1})$ is further in time than $s(c_i)$ exists, such that :

$$q = s(c), \tag{2.28}$$

where q is a shorthand for the subordinator $s(c)$, the price process can then be summarized as:

$$P_{cal}(c) = W\big(s(c)\big). \tag{2.29}$$

In Equation (2.29), the price series $P_{cal}(c)$, is said to be subordinated to the parent process $W\big(s(c)\big)$ and the subordinator $s(c)$ is a càdlàg process that measures market's intrinsic time which flows at variable rates (Velasco-Fuentes and Ng (2010)).

Alternatively, the return series, $r_{cal}(c)$ can be expressed as:

$$r_{cal}(c) = \Delta W\big(s(c)\big), \tag{2.30}$$

where $\Delta W\big(s(c_i)\big) = W\big(s(c_i)\big) - W\big(s(c_{i-1})\big)$.

Sampling under tick time, where t represents transaction time, asset returns, $r_{tick}(t)$ can then be expressed as:

$$r_{tick}(t) = \Delta W\big(s(t)\big) \tag{2.31}$$

where $\Delta W\big(s(t_i)\big) = W\big(s(t_i)\big) - W\big(s(t_{i-1})\big)$.

Then, given that subordinated parent process $\Delta W\big(s(t)\big)$ in Equation (2.31) is a Brownian Motion, normally distributed returns should be obtained by using the transformation:

$$R_{tick}(t) = \frac{r_{tick}(t)}{\sqrt{s(t)}} \sim N(\mu_{tick}, \sigma^2_{tick}), \tag{2.32}$$

where μ_{tick} and σ^2_{tick} are the mean and the variance of the subordinated tick time series, $R_{tick}(t)$ and $r_{tick}(t)$ represent time deformed and raw returns respectively, and $s(t)$ is the subordination vector. All variables are sampled under tick time.

2.3.2 Subordinators

Let "natural time" be defined as the unique subordinator $s_N(t)$, with which the return series achieve perfect "normality" under tick-time sampling. Then the goal of this chapter is to find the best approximation for natural time via the choice of sampling frequency and subordinator $s(t)$, using various linear and nonlinear combinations of volume, duration and order book imbalance parameters.

The linear subordinator utilized in this study can be summarized as:

$$s(t) = \beta X(t) \tag{2.33}$$

where $X(t)$ is the vector of variables[6] that is used to form the subordinator $s(t)$ and β is corresponding vector of coefficients for the variables in $X(t)$.

[6] For transaction sampling sizes larger than 1 tick, $X(t)$ variables are computed by taking into account all available information at each transaction.

Consequently, one of the major contributions of this chapter is finding the vector of variables, $X(t)$, that can be used to achieve subordinated normal returns. The exact forms of the subordinations function are presented in Equations (2.34) – (2.37).

The linear subordinator is of the form:

$$s(t) = \beta_2 \, volume(t) + \beta_3 \, duration(t) + \beta_4 \, Init\ Imb^2(t) + \beta_5 \, Vol\ Imb^2(t). \qquad (2.34)$$

However, to better assess the value of proposed subordinators, additional structural changes to the subordinator function itself was made. An asymmetric subordination function is formed to check for possible differences in the behavior of the subordinator to the sign of returns.

The returns and their corresponding subordinators are classified according to the sign of returns. The positive and negative return series are then used to estimate the coefficients for the subordinators. The corresponding results are combined with the two original return series, classified according to the sign of returns, to produce the subordinated return distribution.

The asymmetric subordinator is of the form[7]:

$$s(t) = \begin{cases} {\beta_2}^+ \, volume^+(t) + {\beta_3}^+ \, duration^+(t) + {\beta_4}^+ \, (Init\ Imb^+)^2(t) + {\beta_5}^+ \, (Vol\ Imb^+)^2(t), & r \geq 0 \\ {\beta_2}^- \, volume^-(t) + {\beta_3}^- \, duration^-(t) + {\beta_4}^- \, (Init\ Imb^-)^2(t) + {\beta_5}^- \, (Vol\ Imb^-)^2(t), & r < 0 \end{cases}. \qquad (2.35)$$

Additionally, given the existing literature on the autoregressive nature of variance, the past values of squared returns were used to augment the subordinator. AR(1) terms are used to test this hypothesis.

[7] The + and – signs indicate the respective series for positive and negative returns.

The autoregressive subordinator function includes past values of the squared returns:

$$s(t) = \beta_1 \, {r_{tick}}^2(t-1) + \beta_2 \, volume(t) + \beta_3 \, duration(t) + \beta_4 \, Init\ Imb^2(t) + \beta_5 \, Vol\ Imb^2(t). \quad (2.36)$$

Finally, the asymmetric and autoregressive models are combined to produce the fourth structural model for the subordinator.

The autoregressive asymmetric subordinator function can be expressed as:

$$s(t) = \begin{cases} {\beta_1}^+ \, ({r_{tick}}^+)^2(t-1) + {\beta_2}^+ \, volume^+(t) + {\beta_3}^+ \, duration^+(t) + {\beta_4}^+ \, (Init\ Imb^+)^2(t) + {\beta_5}^+ \, (Vol\ Imb^+)^2(t), & r \geq 0 \\ {\beta_1}^- \, ({r_{tick}}^-)^2(t-1) + {\beta_2}^- \, volume^-(t) + {\beta_3}^- \, duration^-(t) + {\beta_4}^- \, (Init\ Imb^-)^2(t) + {\beta_5}^- \, (Vol\ Imb^-)^2(t), & r < 0 \end{cases}. \quad (2.37)$$

Subordination essentially aims to account for the heteroscedasticity in asset returns, by utilizing volatility related information. Thus, in many respects, subordination could be classified as a volatility-based approach. The use of past square returns then naturally brings to mind the GARCH model (Bollerslev (1986)). Hence, to make an accurate comparison, a GARCH(1,1) model is separately estimated. Returns are then subordinated using these estimated GARCH parameters to construct a benchmark model.

The GARCH(1,1) model used in estimations can be summarized as follows:

Let error term ϵ_{tick} represent the mean-adjusted returns, which can be decomposed into a time-varying standard deviation σ_{tick} and a stochastic component $Z_{tick} \sim N(0,1)$.

$$\epsilon_{tick}(t) = \sigma_{tick}(t) \, Z_{tick}(t). \quad (2.38)$$

Then the conditional variance under a GARCH(1,1) specification can be expressed as:

$${\sigma_{tick}}^2(t) = \varphi_0 + \varphi_1 {\epsilon_{tick}}^2(t-1) + \omega_1 {\sigma_{tick}}^2(t-1), \quad (2.39)$$

where $\varphi_0 > 0$, $\varphi_1 \geq 0$, $\omega_1 \geq 0$ and $\varphi_1 + \omega_1 < 1$.

2.3.3 Maximum Likelihood Estimation

Maximum likelihood estimation (MLE) methodology is used to estimate the coefficient vector β where a (subordination-adjusted) normal distribution is specified as the resulting distribution. The subordination-adjusted log likelihood function that is employed in MLE estimations takes into account the fact that this subordinated return series follow a normal distribution with unknown but finite mean and variance. Additionally, the use of 100-tick sampling frequency dampens autocorrelations within the tick data.

Given this structure, the joint probability distribution function for the subordinated tick time series can be expressed as:

$$f\left(R_{t_1}, R_{t_2}, \dots, R_{t_n} \middle| s(t), \mu_{tick}, \sigma_{tick}^2\right) \tag{2.40}$$

where R_{t_i} represent elements of the time deformed return series, $R_{tick}(t)$, described in Equation (2.32).

Equation (2.40) can also be expressed as:

$$f\left(R_{t_1}, R_{t_2}, \dots, R_{t_n} \middle| s(t), \mu_{tick}, \sigma_{tick}\right) = \frac{1}{\sigma_{tick}{}^n\left(\sqrt{2\pi}\right)^n} \, exp\left\{-\frac{1}{2}\sum_{tick=1}^{\infty}\frac{\left(\frac{r_{tick}(t)}{\sqrt{s(t)}}-\mu_{tick}\right)^2}{\sigma_{tick}^2}\right\}. \tag{2.41}$$

Then the log-likelihood function is:

$$\ln LF\ (s(t), \mu_{tick}, \sigma_{tick}) = -\frac{n}{2}\ln\sigma_{tick}^2 - \frac{n}{2}\ln 2\pi - \frac{1}{2}\sum_{tick=1}^{\infty}\frac{\left(\frac{r_{tick}(t)}{\sqrt{s(t)}}-\mu_{tick}\right)^2}{\sigma_{tick}^2}. \tag{2.42}$$

Similarly, the log-likelihood function for GARCH(1,1) estimation is:

$$\ln LF\ (\mu_{tick}, \sigma_{tick}) = \sum_{i=1}^{n}\left(-\frac{1}{2}\ln 2\pi - \frac{1}{2}\sigma_{tick}{}^2(t_i) - \frac{1}{2}\frac{\epsilon_{tick}{}^2(t_i)}{\sigma_{tick}{}^2(t_i)}\right). \tag{2.43}$$

2.3.4 Evaluation

To evaluate the ability of the linear subordinators to transform the tick returns into a normally distributed series, time deformed return series are tested with Kolmogorov-Smirnov (KS) and Jarque-Bera (JB) tests. The use of KS test is justified by the large number of observations present in the dataset. Additionally, unlike its successor Anderson-Darling test or the Jarque-Bera statistics, KS test is known to be sensitive to the location parameter due to its focus on the maximum difference between two distributions. Given this setup, optimization procedure for the KS test can be expressed as:

$$min\, KS\ (\alpha, R_{tick}(t)). \tag{2.44}$$

The Kolmogorov-Smirnov statistic KS shown in Equation (2.44) is calculated as:

$$KS = sup\ |F(R_{tick}(t)) - F_G(R)| \tag{2.45}$$

where $F(R_{tick}(t))$ is the empirical distribution function and $F(R)$ is the Gaussian cumulative distribution function.

The JB test statistic measures the deviation from normality in skewness and kurtosis parameters, a fit choice for financial return series as they exhibit most severe deviations from normality in their higher moments. However, unlike Velasco-Fuentes and Ng (2010), JB test is used to validate the results of the MLE procedure rather than estimate the coefficients for the parameters used in the subordinator.

The Jarque-Bera statistic can be calculated as:

$$JB = \frac{n}{6}\left(Skew^2 + \frac{1}{4}(Kurt - 3)^2\right), \tag{2.46}$$

where n is the number of observations, $Skew$ and $Kurt$ are sample skewness and kurtosis respectively.

2.4 Data & Analysis

The high frequency dataset utilized in this chapter uses Level 2 SETS data from the London Stock Exchange (LSE), where stocks are traded in a continuous-time double auction system. The LSE sorts and matches orders first by their price competitiveness and then by their time of submission. The Level 2 dataset includes the whole order book depth at any given point in time as well as the actual trade times and prices for realized trades. The order book data includes "public" orders that appear on the order book and excludes order types such as non-persistent or Iceberg orders. Hence, the bulk of the information contained in the order book stems from limit and market orders[8].

The period under study spans from July 2007 to June 2008. Taking into account the large market swings during this time, the whole dataset is split into four 3-month periods, where the first period (P1) spans from July 2007 to September 2007. Similarly, P2 covers October 2007 – December 2007, P3 January 2008 – March 2008 and P4 April 2008 – June 2008. Top ten stocks with highest liquidity are selected for the purpose of this study[9]. Each stock is analyzed on a period by period basis so as to not include irrelevant past data. This partitioning of data is warranted by the wild swings that dominated financial markets during the sample period.

[8] The details of the order book reconstruction can be found in Appendix A.

[9] The list of stocks used is presented in Appendix B.

2.4.1 Sampling Frequency

The sampling frequency, whether one is using calendar or transaction time, has a substantial impact on the raw returns one observes. Hence, in order to determine an optimal frequency, the effects of sampling frequency on autocorrelation and the distribution of the returns were examined.

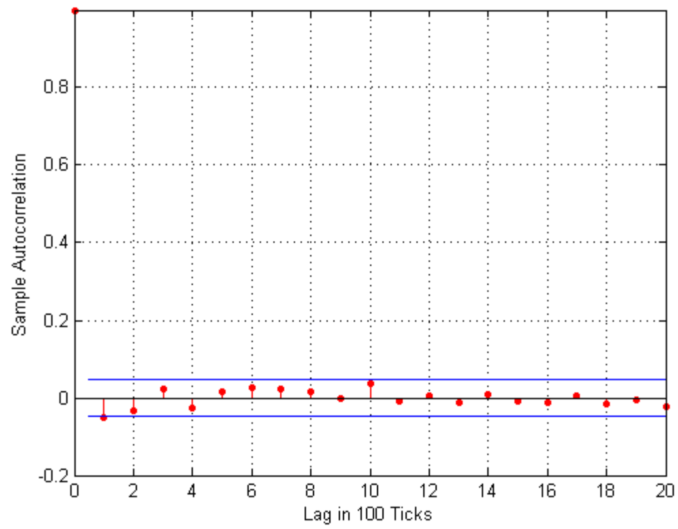
2.4.1.1 Sampling Effects on Autocorrelation

The first obstacle one needs to address when working with financial series is autocorrelation, as it may undermine the inferences made. This phenomenon becomes even worse as the sampling frequency is increased. The fourth period for HSBC stock was chosen for exemplification purposes and Ljung-Box test was applied to several sampling tick sizes using a lag size of 20. Autocorrelation was present up to a sampling frequency of 100 ticks. Autocorrelation and partial autocorrelation functions for HSBC P4 with a sampling frequency of 100 ticks were also mapped via a correlogram and ACF and PACF decay rate did not converge albeit being small. Similar results were obtained for other stocks. ACF functions and Ljung-Box (LB) test statistics for HSBC are presented in Table 2.1 and Figure 2.3 respectively.

Table 2.1: Ljung-Box Test (Lag=20)

Sampling Tick	LB Test	p-value
5	226.8479	0
10	105.6981	< 0.0001
20	56.3034	< 0.0001
50	33.5694	0.0292
100	16.1198	0.7092
200	17.0837	0.6475

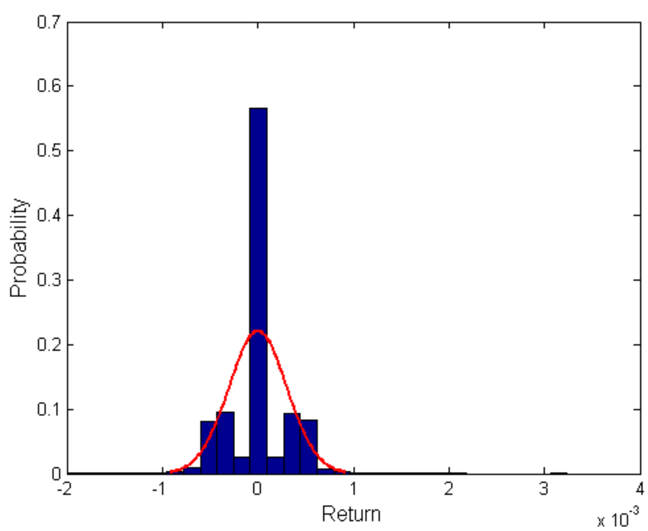
Figure 2.3: Autocorrelation Function for Returns of HSBC Stock Prices in Period 4 Sampled at 100 Ticks



2.4.1.2 Sampling Effects on Distribution of Returns

Due to the nature of ultra-high frequency data, additional measures to deal with price discreteness were necessary. Figure 2.4 shows the return histogram fitted on a normal distribution curve for tick returns.

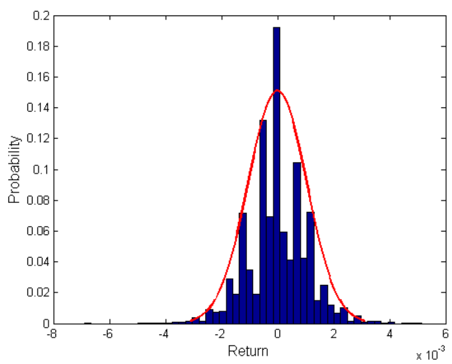
Figure 2.4: Histogram for Tick Returns for HSBC Stock Prices in Period 4



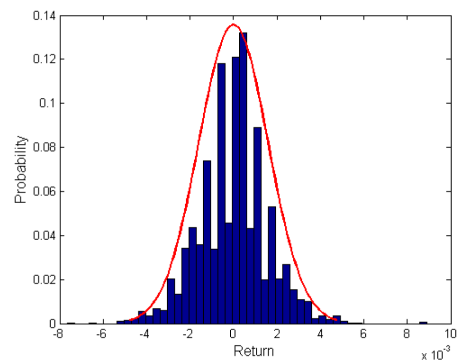
As is apparent from Figure 2.4, raw returns at the single-tick sampling frequency are dominated by price discreteness. Hence, several sampling frequencies were tested to ascertain the exact effects of sparse sampling on the distribution of returns. The graphs in Figure 2.5 illustrate the relationship between decreasing sampling frequency and return distribution.

Figure 2.5: Distribution vs. Sampling Frequency: HSBC Stock Returns in Period 4

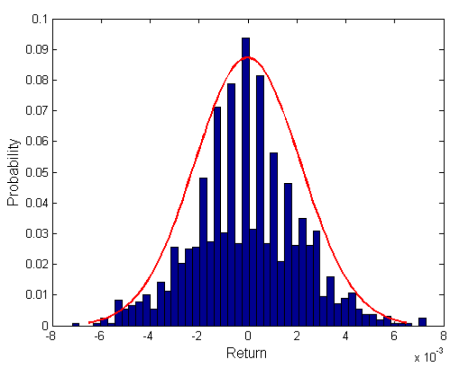
Sampling Frequency: 20 Ticks



Sampling Frequency: 50 Ticks



Sampling Frequency: 100 Ticks



Sampling Frequency: 200 Ticks

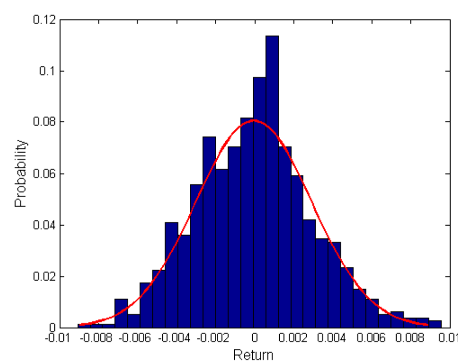


Figure 2.5 shows that at sparser sampling frequencies[10] the return distribution approaches normality. However, simple aggregation of returns to produce normality is neither new to

---

[10] The sampling frequencies shown throughout Chapter 2 use non-overlapping sections of the data. Hence, at no point in time past information is used twice in this analysis.

the literature, nor would it be feasible in a subordination study which tests the limits of the sampling frequency under which subordination still produces normality. Thus, to assess the exact effects of sampling frequency on price discreteness and the distribution of returns and to determine an optimal sampling frequency for natural time, the first four moments are computed. Table 2.2 contains the results.

Table 2.2: Sampling Frequency vs. Moments: HSBC Stock Returns in Period 4

Sampling Frequency	Mean	Variance	Skewness	Kurtosis
1-Tick	-2.41 e-7	9.34 e-8	0.0154	4.4958
5 Ticks	-1.30 e-6	3.09 e-7	0.0103	4.1357
10 Ticks	-2.67 e-6	5.77 e-7	0.0636	4.1606
20 Ticks	-3.71 e-6	1.08 e-6	0.0345	4.2115
30 Ticks	-6.97 e-6	1.55 e-6	0.0672	4.1877
40 Ticks	-6.12 e-6	2.07 e-6	0.0327	4.1762
50 Ticks	-4.70 e-6	2.52 e-6	0.0488	4.0486
60 Ticks	-9.25 e-6	2.93 e-6	0.0989	4.0019
70 Ticks	-6.56 e-6	3.45 e-6	-0.0017	3.6552
80 Ticks	-9.57 e-7	3.83 e-6	0.0067	3.3744
90 Ticks	-1.12 e-5	4.37 e-6	0.0056	3.6115
100 Ticks	-1.37 e-5	4.67 e-6	-0.0020	3.2000
200 Ticks	-1.86 e-5	5.96 e-6	0.0108	3.2685
300 Ticks	-3.24 e-5	6.94 e-6	-0.0046	3.2953
400 Ticks	4.17 e-6	7.93 e-6	0.0663	3.0219
500 Ticks	-4.66 e-5	8.85 e-6	0.1828	3.1304

Table 2.2 suggests the use of 100 ticks as the sampling frequency is appropriate, as sampling at lower frequencies after 500 ticks causes further negative skewness and abnormally low kurtosis values for a high frequency return series. The results presented in Table 2.2 were reproduced for all stocks and periods but they are not included here to conserve space. However, the effects of sampling frequency do not vary much from stock to stock. Thus, a sampling frequency of 100 ticks is used for all stocks and periods unless mentioned otherwise. In cases where different sampling frequencies have been used, the moments of the resulting raw distribution were utilized to determine the new sampling frequency[11]. A sampling frequency of 100 ticks generally resulted in 1,500 data points per period.

2.4.2 Subordination Variables

Upon selection of the sampling frequency, the influential variables discussed in the previous sections can now be tested for validity. Trade volume, cumulated across the selected number ticks, and its log transformation are used to find the impact of trade size on price formation. Duration between each sampling point is also used to assess the urgency with which orders have been filled. In order to assess how the liquidity state of the market influences price movements, the imbalance in the order book is computed in various different ways. The *Imbalance* term cumulates the volume difference between bid and ask sides for the whole depth of the order book and averages this number for across

[11] Figure 2.5 shows that at a sampling frequency of 100 ticks, returns are not normally distributed. Before each subordination procedure, the raw distribution of returns is checked and the sampling frequency is increased if raw returns are normally distributed.

the selected sampling frequency. Similarly, $Level\ 1\ Imbalance$ and $Level\ 3\ Imbalance$ apply the same procedure to the first 1 and 3 levels from the top of the order book, respectively.

The number of transactions has been previously used by Ané and Geman (2000) to subordinate the price processes. This measure provides partial information on the number of entities involved, but does not make any distinction between the direction of trades. Thus, a more transparent measure is needed, which can be obtained by looking at the difference in the number of unique trades in a given interval. At each tick, which may include multiple buy and sell orders, the number of initiators for each side is found and the difference is recorded. This number is then cumulated for the span of sampling frequency and divided by the number ticks to form $Initiator\ Imbalance$ variable. The same process is repeated for $Volume\ Imbalance$ taking into account the volume of trades. A negative number means excess sell side orders, where as a positive number denotes buy side for these two variables. Finally, log transformations of squared $Initiator\ Imbalance$ and $Volume\ Imbalance$ are added into the list of possible variables. Although the squared order book variables lose information on whether it was the buy or the sell side orders that were in excess, this transformation is dictated by the non-negativity constraint presented in Equation (2.32). Additionally, squared order book variables are expected to be a better gauge for market volatility given their construction.

Table 2.3: Regression Analysis[12]: Mean Adjusted Squared Returns for HSBC in P4 Sampled at 100 Ticks

Subordinator	**Regression Statistics**		
	constant	**coefficients**	**R^2**
Volume	1.8288 e-6	2.5488 e-12 (0)	0.0211
Duration (sec)	6.0325 e-6	-1.2757 e-9 (0)	0.0099
Imbalance	4.6929 e-6	-7.3210 e-14 (0.8409)	0
Level 1 Imbalance	4.6660 e-6	6.5167 e-12 (0.5088)	0.0003
Level 3 Imbalance	4.6626 e-6	3.6090 e-12 (0.2688)	0.0007
Initiator Imbalance	4.6872 e-6	-2.8378 e-7 (0.3850)	0.0004
Volume Imbalance	4.6357 e-6	7.1408 e-11 (0.2607)	0.0008
Log-Volume	-3.7174 e-5	3.0175 e-6 (0)	0.0221
Log-InitImb^2	4.0356 e-6	3.0759 e-8 (0)	0.0121
Log-VolImb^2	-1.8930 e-6	4.6257 e-7 (0)	0.0251

Table 2.3 shows a peculiar outcome. None of the standing order book variables that describe market liquidity conditions, namely $Imbalance$, $Level\ 1\ Imbalance$ and $Level\ 3\ Imbalance$ are found to be significant in explaining squared returns. This is an unexpected finding, which suggests that variables related to the active trading environment already contain the necessary liquidity information. For this reason, all standing order book variables are dropped from further study.

[12] The values in parentheses in Table 2.3 and all of the tables that follow show respective p-values for each variable.

Additionally, *Initiator Imbalance* and *Volume Imbalance* are also removed from further analysis, as per the non-negativity constraint[13]. Although the remaining five subordinators are significant in normalizing the return series at the 5% significance level, confirming the findings of Clark (1973) and Ané and Geman (2000), volume is also dropped from further subordination runs as similar results can be produced by the log-volume.

2.5 Results

The subordination methodology employed in this chapter entails maximum likelihood estimation. As with any optimization problem, local minima and maxima may constitute a problem. Although this was not the case in this study, I start with reporting single run results for subordination and then move on to the global procedure used for the whole of the dataset to overcome any possible local extrema problems.

2.5.1 Subordination Results: Single Run

The single run multiple subordination results presented in this section use a single starting point to estimate the coefficients for the subordinator. The coefficients, p-values, log-likelihood function value as well as KS and JB test statistics for the subordinated returns using the produce described in Equations (2.34)-(2.37) are presented in Table 2.4:

[13] Logarithms of squared initiator and volume imbalance are referred to as initiator imbalance and volume imbalance from this point on.

Table 2.4: Multiple Subordination[14] Results for HSBC Returns in P4 Sampled at 100 Ticks

Single Run

Subordinator	*Linear*	*Autoregressive*	*Asymmetric*[15]		*Autoregressive Asymmetric*	
μ	1.2869 e-14 (1)	-9.7291 e-11 (0.1538)	-3.5890 e-10 (0.0332)		1.4613 e-11 (0.0591)	
σ	2.6896 e-10 (1)	6.3386 e-10 (0)	7.9000 e-9 (0)		3.1514 e-10 (0)	
r_{tick-1}^2	-	7.3386 e+4 (0)	-	-	3.7485 e+5 (0)	2.5673 e+5 (0)
Volume	1.9926 e+6 (0)	1.7613 e+5 (0)	1.7673 e+3 (0)	1.7660 e+3 (0)	9.2414 e+5 (0)	7.0897 e+5 (0)
Duration	2.4668 e+6 (0)	1.8243 e+5 (0)	1.7373 e+3 (0)	1.7701 e+3 (0)	9.6022 e+5 (0)	7.4304 e+5 (0)
Log-InitImb^2	7.5385 e+4 (0)	1.2095 e+3 (0)	2.4497 e+1 (0)	2.5722 e+1 (0)	6.6751 e+3 (0)	4.7491 e+3 (0)
Log-VolImb^2	1.9315 e+6 (0)	1.7633 e+5 (0)	1.4189 e+3 (0)	1.4644 e+3 (0)	9.0276 e+5 (0)	7.3118 e+5 (0)
Log-likelihood	-11,803	-9,904	-6,034		-11,203	
KS Test	0.0437 (0.0031)	0.0442 (0.0026)	0.0474 (9.9740 e-4)		0.0443 (0.0026)	
JB Test	18 (0.0010)	15 (0.0018)	52 (0.0010)		16 (0.0014)	

[14] The subordination results reported here and henceforth multiplies tick returns with 1e+6 and divides $Log - Initimb^2$ term by 100 as not to compromise floating point calculations in Matlab. Likewise, results for the duration term are reported for duration measured in minutes.

[15] For the autoregressive and asymmetric autoregressive models, estimated mean and standard deviation is the same with respect to the sign of returns since subordinated returns are expected to come from a single normal distribution. Separate coefficient estimates has been reported for other variables in each column. The estimates for positive returns can be found on the left hand side while the estimates for negative returns are reported on the right hand side.

The multiple subordination results presented in Table 2.4 points to a striking conclusion: neither asymmetric or autoregressive asymmetric models produce significantly different results from the remaining models. The functional values for asymmetric or autoregressive asymmetric models and their corresponding p-values for both KS and JB tests are no better than those obtained with the linear or autoregressive approaches. The results presented in Table 2.4 extend to other stocks and periods.[16] Contrary to the asymmetric approach, the autoregressive model is found to augment the linear model, further supporting the use of past squared returns. Moreover, the significance of imbalance terms in addition to volume and duration parameters seems to solidify the notion that order book information is important in subordination, hence variance estimation.

Another interesting finding present in Table 2.4 is that the coefficient for the duration term is positive. Although one would expect volatility to be high during rapid trading periods, the findings point to the opposite. This is possibly due to the use of transaction time sampling, which inherently accounts for the market's intrinsic time. For example, if one were to look at the market opening hours, the high frequency of trades would trigger numerous sampling points within a short time interval. Thus, for a given sampling frequency, although many trades would come to pass, one would not observe a substantial price change in the value of the asset. By the same token, one would also observe larger price changes during the rest of the trading day using the same tick sampling frequency, as the time between trades would be substantially larger.

[16] Asymmetric approaches are omitted from further reporting as findings extend to other stocks.

2.5.2 Subordination Results: Global Procedure

The single run multiple subordination The findings presented in Table 2.4 may however be subject to the ubiquitous local extrema problem as the findings are produced on a single-run. To address this possible shortcoming, the gradient-based optimization algorithm is augmented with 10^5 different starting points to cover a vast search space.[17] The results for HSBC stock in each period using this procedure (Global) are reported in Table 2.5. Further details of subordination results for all stocks and periods are presented in Appendix C.

Table 2.5: Multiple Subordination Results using Global Procedure

HSBC

Normality		**P1** (100Ticks)		**P2** (100Ticks)		**P3** (100Ticks)		**P4** (100Ticks)	
		Linear	*Autoregressive*	*Linear*	*Autoregressive*	*Linear*	*Autoregressive*	*Linear*	*Autoregressive*
KS Test		0.0748 (1.2026 e-7)	0.0663 (4.2692 e-6)	0.0474 (8.2207 e-4)	0.0481 (6.5517 e-4)	0.0503 (2.4969 e-5)	0.0477 (7.7809 e-5)	0.0456 (0.0018)	0.0448 (0.0022)
JB Test		77 (0.0010)	139 (0.0010)	642 (0.0010)	238 (0.0010)	554 (0.0010)	1,248 (0.0010)	2 (0.2947)	0 (0.5000)
GARCH	KS Test	0.0529 (4.9487 e-4)		0.0539 (8.3033 e-5)		0.0395 (0.0019)		0.0415 (0.0059)	
	JB Test	6 (0.0430)		59 (0.0010)		57 (0.0010)		2 (0.4278)	

Table 2.5 shows that findings regarding asymmetric subordination of HSBC in P4 using single-run method can be extended to all periods and stocks. One possible reason for the failure of asymmetric models could be an inherent stability of information flow through the selected subordinators. Hence, it can be argued that volume, duration, initiator imbalance and volume imbalance variables effect returns much in the same way whether the market is moving upward or downward. As such, asymmetric models could not produce superior results by treating returns of opposite signs differently.

[17] All further results reported use 10^5 starting points.

Furthermore, as is apparent from Table 2.2, the choice of sampling frequency, which constitutes an important part of the natural time approach, has a dominant effect on the distribution of raw returns. While sparse sampling mitigates price discreteness, it eventually reduces the relevance of past order book data. For this reason, a sampling frequency of 100 ticks was used for all stocks except for second and fourth periods of SAB Miller. In these periods, normally distributed returns were obtained without the need for subordination at 100 ticks. Hence, higher sampling frequencies were chosen to produce comparable raw distributions in terms of their first four moments.

Closer examination of the results in Appendix C reveals an interchangeability between log-volume and volume imbalance terms.[18] Either one of two subordinators, when used in conjunction with others, is significant but they fail to be significant together on several occasions. While volume imbalance is significant for SAB Miller and Shell in the second period, the reverse holds for HSBC. In contrast, both variables are significant for all periods for Vodafone. Nonetheless, a combination of volume and initiator imbalance seems to be the better choice in general. This interchangeability can be caused by the structural changes in the way variance related information is conveyed in the market. It might be the case that in some periods, a combination of volume and initiator imbalance captures variance related information while in others volume imbalance proves to be a better gauge.

[18] Further details of the subordination results can be found in Appendix C.

Furthermore, convergence of volume and volume imbalance terms, which would convey similar information when orders are one-sided, can also render the volume imbalance term redundant. One or both of these factors may be at work in a given period as they are by no means mutually exclusive.

The autoregressive subordination model, which uses past squared returns, was also found to perform generally better than the linear model for all stocks. Although similar results could be obtained using the linear model in several of the periods where normally distributed subordinated returns were produced with the autoregressive model, this was not possible for the second period of SAB Miller and Diageo and third period of Shell.

In comparison with autoregressive subordination, GARCH(1,1) model does a marginally better job in periods where subordination fails to produce normally distributed returns. However, in cases where normally distributed returns were obtained via subordination, GARCH not only produced worse results but also failed to achieve normality with the exception of three instances, second and fourth periods of British American Tobacco and first period of BG Group. These three cases where GARCH produced better results compared to subordination could very well be due to a local minima problem in the subordination procedure.

All in all, for a total of forty periods and 10 stocks, subordination resulted in normally distributed returns in nine periods, while GARCH based subordination could only produce normal returns in the five periods.[19]

[19] The resulting distributions from multiple subordination and GARCH(1,1) were assumed to be normally distributed, if they have passed either one of the KS or JB tests.

In two of these five periods where GARCH based subordination was successful, normality was also achieved with linear and autoregressive subordination. On the other hand, none of the linear, autoregressive or GARCH subordination methodologies could produce normally distributed returns for British Petroleum, GlaxoSmith Kline and Rio Tinto in any period.

The results presented in this section diverge from the existing subordination literature in many fronts. First of all, one of the most turbulent periods for single stocks was examined in this chapter which renders the effort to produce normally distributed returns inherently much harder.

Both subordination studies that utilized single stocks, Ané and Geman (2000); Silva and Yakvenko (2007), use data dating before year 2000. In contrast, I use a much recent dataset which reflects the conditions of today's financial markets better. Clark (1973) and Velasco-Fuentes and Ng (2010), on the other hand focused on cotton and FTSE-100 futures that are not even subject to the idiosyncrasies of single stocks. Furthermore, unlike extant studies in the literature, this chapter samples the data in tick time while presenting the results for an unmatched total number of stocks, periods and models, including a GARCH based subordination procedure as well.

2.6 Discussion of Results

The work presented in this chapter focuses on the application of stochastic subordination to high-frequency returns sampled under transaction time. Several previous subordination based studies have been performed using calendar time (Clark (1973); Ané and Geman (2000); Silva and Yakvenko (2007); Velasco-Fuentes and Ng (2010)). Furthermore, only a subset of the variables used in this research were employed in the above mentioned studies. Order book variables, which contain information on both market liquidity and the initiator of trades, have been added into the subordination procedure, which is another novel contribution of this paper to the literature. This subordination procedure, which operates under tick time and uses order book variables to transform the return series into a normally distributed one, is referred to as "natural time" in this chapter.

Previous studies have found volume and number of trades to carry relevant information to price formation under physical time (Clark (1973); Ané and Geman (2000); Silva and Yakovenko (2007); Velasco-Fuentes and Ng (2010); Huth and Abergel (2012)). Their counterparts in transaction time, volume and duration, are also found to be significant in stochastic subordination. The results show that order book variables and past squared returns also carry important variance-related information. The addition of these variables into the subordinator augments the model such that subordinated returns are normally distributed in most cases.

The GARCH terms for the exogenous GARCH(1,1) model with the order book variables, were insignificant in all periods for all stocks. This consistent superiority of natural time approach to the benchmark GARCH model has profound implications. The success of the natural time approach not only supports the normal distribution assumption but also indicates that transaction time might be the right sampling methodology when using high frequency data. Furthermore, as the ability to successfully normalize returns via subordination essentially hinges on accounting for heteroscedasticity, the order book variables used to subordinate returns can also be used to forecast volatility given the clear information advantage they provide over GARCH.

The research in this chapter introduces a novel way to view financial returns while also giving the reader a set of possible variables that are effective in accounting for volatility. In this respect, Chapter 2 not only fulfills the first goal of this thesis by achieving normally distributed subordinated returns but also offers an unconventional volatility forecasting strategy. Chapter 2 makes three major contributions to the literature. First, by successfully recovering normality of high frequency returns via subordination, this chapter presents evidence in support of the normal distribution assumption behind numerous finance theories. Second, by successfully achieving subordinated normal distributions, Chapter 2 also demonstrates that transaction time sampling is a better alternative to calendar sampling, especially when using high frequency data. Third, by creating a volatility gauge from order book variables, Chapter 2 also contributes to the volatility literature by providing evidence for order book based volatility forecasting methodologies.

Market players that have access to the type of order book data used in this chapter may be able to foretell imminent excess volatility episodes and adjust their positions and leverage accordingly. Financial authorities which oversee stock markets could also use the information contained within the order book to prevent a disorderly collapse of the system. Either use of this information will contribute to the efficiency of financial markets.

In Chapter 3, I build upon the findings of Chapter 2 and test the validity of the influential variables used in natural time approach as crash predictors. Using both futures and single stock data, I combine these variables under a linear discriminant analysis framework to successfully forecast high frequency crashes. To follow, Chapter 4 focuses on macroeconomic crashes which manifest themselves as currency devaluations. Several binary and panel models are tested and results indicate that successful crash prediction is possible.

Chapter 3: Flash Crash, Liquidity Dynamics and Market Heat

3.1 Introduction

The primary focus of this chapter is to predict flash crashes, sudden price movements in high frequency returns. While the scope and the methods employed in this chapter are dissimilar to the ones used in the previous chapter, Chapter 3 builds upon the findings of Chapter 2. Given the ability of order book variables to successfully subordinate returns and achieve normality, the next logical step is to use these variables to predict high frequency crashes. To this effect, I combine the variables found to be significant in Chapter 2 with linear discriminant analysis to predict flash crashes in two very different financial markets, namely E-Mini S&P500 futures and selected LSE stocks. Furthermore, contrary to imperfect classification methodologies used in all existing flash crash literature, high frequency trades are classified perfectly into buyer or seller initiated trades in this research.

"Market heat" introduced in this chapter is a prediction tool for sudden asset price depreciations. It measures the increased activity in the order book in order to make its predictions. Much like the way we measure the excited movements of particles as "heat" in thermodynamics, market heat measures the state of urgency in the market via the use of order book data. Market heat is based on market microstructure literature and accounts for the short term dynamics observed in today's financial markets. Market heat outperforms all alternatives tested here and successfully predicts flash crashes in both markets studied. Market heat's 5-minute ahead forecasts provide ample supply of time to stock exchanges and investors to protect themselves against impending price fluctuations. In this respect, Chapter 3 not only contributes to the growing flash crash literature by presenting a solid method to predict high frequency crashes but also offers stock exchanges a potential circuit breaker to avoid future flash crashes. Hence, Chapter 3 covers part of the second goal of this thesis, namely, prediction of financial crashes with varying time horizons.

Despite their infrequent occurrence the financial arena is plagued by crashes that erase years' worth of capital earnings. The notorious stock market crash of October 1929 is perhaps the best known among these, marking the beginning of the Great Depression despite joint efforts to sustain the stock market and the economy. Similar significant stock market crashes include the Black Monday (October 19th, 1987) caused by the "soft-landing" of the U.S. economy and the Black Wednesday (September 16th, 1992) where sterling pound was forced out of the European Exchange Rate Mechanism. One common characteristic of all of the above described crashes, except for the fact that they all occurred in autumn, is that the financial losses incurred could only be recovered after several years.

Increasingly the changes in the structure of global markets usher in a new breed of financial crashes, namely flash crashes. The mini crash of October 27th, 1997 where the market pared over 60% of its losses the following day or The Flash Crash of May 6th, 2010 are ostensibly acute forms of these flash crashes in which sudden order imbalances cause abrupt price changes.

In this chapter, the predictability of sudden price dips in asset prices is investigated and a new metric to signal impending crashes is proposed. The proposed metric, "market heat" takes a liquidity based approach to forecasting mini crashes since as in a high frequency setting it is often the liquidity conditions rather than sudden changes in the fundamentals that dictate price moves. As shown in Ng (2008), additional time costs are attached to block trades as financial markets are unable to absorb large orders in short time intervals. Market heat (MH) takes into account not only the order imbalances, but also the amount of liquidity available and the speed with which trades are being initiated. By combining these three important elements in a nonlinear fashion, MH tackles the elusive problem of crash prediction.

To the best of my knowledge, a nonlinear signaling metric which explicitly includes liquidity to predict mini crashes has not been suggested before. Using tick data for E-Mini S&P500 futures and LSE stocks, I test MH against a linear and a time-bucketed market microstructure based metric - similar to the one introduced in Easley, Lopez de Prado and O'Hara (2012). MH outperforms its counterparts in both markets across a set of binary classification measures. The robustness of results across markets and different time frames supports the case for a nonlinear liquidity based approach to crash prediction. In addition to the general success of MH, the ability of all three metrics used here to capture the Flash Crash underscores a simple fact: the Flash Crash of May 6th 2010 could have been avoided if MH metric proposed in this study were employed by the Chicago Mercantile Exchange (CME) as a circuit breaker.

Section 3.2 reviews the relevant high frequency crash literature and introduces findings of previous work on the Flash Crash. In Section 3.3, the methodology used to construct MH is explained while Section 3.4 details the two different datasets used. The results are presented in Section 3.5 and Section 3.6 concludes.

3.2 Literature Review

Similar to the case of program trading becoming the culprit for Black Monday, high frequency trading instantly became the culprit for the sudden dip in the E-Mini S&P 500 futures during the Flash Crash. However, the findings of the 2010 SEC Report and Kirilenko et al. (2014) regarding the Flash Crash suggest otherwise. Kirilenko et al. (2014) report the main cause to be the automated execution of a large sell order by a fundamental trader 20 minutes prior to the crash, which drained the market liquidity and forced a number of liquidity providers out of the market.

Kirilenko et al. (2014) define 6 market participant categories in their study, namely high frequency traders, intermediaries, fundamental buyers, fundamental sellers, small traders and opportunistic traders. Within this setup, they found high frequency traders' positions to be not large enough to induce the dramatic movements of May 6th. However, this result ties directly with the way Kirilenko et al. (2014) define high frequency traders and may need further investigation. Additionally, the stop loss orders of several liquidity providers combined with the reversal of long high frequency traders' positions at the outset of the crash which removed liquidity from the market are found to exacerbate the fall on May 6th.

Easley, Lopez de Prado and O'Hara (2012) examine the behavior of order imbalances before the crash and develop a volume based flow toxicity measure to make inferences about an impending flash crash, which they dub "Volume-Synchronized Probability of Informed Trading" or VPIN. Easley, Lopez de Prado and O'Hara (2011a, 2011b) compare VPIN to VIX and argue a tradable VPIN contract (FVPIN) could have allowed market makers to condition themselves better against the order flow. The VPIN measure is closely related to the microstructure model of Easley and O'Hara (1992). However, despite its theoretical background, the VPIN measure suffers from artificially introduced errors. The bulk classification methodology employed in Easley, Lopez de Prado and O'Hara (2012, 2015) first cumulates trades within a given time bar and then classify them as buys or sells using a normal or Student's t-distribution respectively. Easley, Lopez de Prado and O'Hara (2015) also argue the classifying trades with the tick rule fails to capture the information within the order flow.

Andersen and Bondarenko (2014a) question the applicability of the VPIN measure and its properties. Andersen and Bondarenko (2014a) use a slightly different data set and compare three different variations of the original VPIN introduced in Easley, Lopez de Prado and O'Hara (2012). The tick-rule VPIN (TR-VPIN) assigns all trades within a given time bar as either buys or sells while bulk volume VPIN (BV-VPIN), which is identical to the original VPIN, assigns trades probabilistically. Andersen and Bondarenko (2014a) also introduce a third measure, namely the fixed bin VPIN (FB-VPIN), which uses volume bars instead of time bars to classify trades. They argue this method is a much more compatible approach given Easley, Lopez de Prado and O'Hara (2012)'s reasoning that financial markets operate under a volume clock. Although the bulk of Andersen and Bondarenko (2014a)'s findings relate to TR-VPIN and FB-VPIN, several findings do carry over to the original VPIN. Andersen and Bondarenko (2014a, 2014b) argue VPIN is highly sensitive to trading intensity and sequencing of trades as such VPIN levels are noticeably affected by the length of the time bars. Andersen and Bondarenko (2014a) also test their VPIN's predictive power and argue that the value of VPIN before the Flash Crash did not provide a clear signal.

Easley, Lopez de Prado and O'Hara (2014) address the concerns raised in Andersen and Bondarenko (2014a) by arguing that bulk classification provides superior results when compared to the tick rule and it is designed to forecast toxic order flow rather than volatility. Easley, Lopez de Prado and O'Hara (2014) and Andersen and Bondarenko (2014b, 2015) provide contrasting studies with respect to the trade classification accuracy. Wu et al. (2013), which uses maximum intermediate return in between two sampling points as a realized volatility measure, find VPIN to be a superior liquidity-induced volatility forecaster "with false positive rates as low as 7%". While Chakrabarty et al. (2013) find the tick rule to be more accurate.

As can be observed from the literature, the focus of many recent studies about the Flash Crash have shifted from developing an actual crash metric that works to determining which method produces a better trade classification accuracy. The advent of detailed high frequency data; however, enables us to easily overcome such trade classification issues.

In this thesis, trades are categorized as bid or ask initiated either by the tags provided by the exchange or by the time of order submission. Hence, unlike any of its predecessors in the flash crash literature, perfect classification is achieved in this chapter, where inferences made about the crash predictors are not subject to any classification error. Perfect classification allows me to capture the true values of these order book variables instead of the distorted approximations we see in the extant literature. Any information carried by the order flow is directly reflected on the variables used for crash prediction.

Thus, the integrity of the data used to form MH and the inferences made with it are beyond reproach; which is another novel contribution of this chapter to the flash crash literature.

3.3 Methodology

The VPIN measure is closely related to market microstructure model of Easley and O'Hara (1992), details of which were presented in Section 2.2.1.1. In this model, the initial spread of equally probable good or bad events is given by:

$$\Psi = \frac{\alpha\upsilon}{\alpha\upsilon+2\xi}\,(P^{+} - P^{-}), \tag{3.1}$$

where Ψ is the bid-ask spread, α is the probability of an information event, υ is the arrival rate of informed trades, ξ is the arrival rate of uninformed trades, P^{+} is the value of the asset given positive news and P^{-} is the value of the asset given negative news.

Similarly the probability of an informed trade (PIN) is defined as:

$$PIN = \frac{\alpha\upsilon}{\alpha\upsilon+2\xi}\,, \tag{3.2}$$

which is the ratio of informed orders to total orders.

As shown in Easley, Engle, O'Hara and Wu (2008), order imbalance can be used as a proxy for informed trading such that VPIN then becomes:

$$VPIN = \frac{\alpha\upsilon}{\alpha\upsilon+2\xi} = \frac{\alpha\upsilon}{V} = \frac{\sum_{b=1}^{n}|V_\tau^S - V_\tau^B|}{nV}, \tag{3.3}$$

V_b^B and V_b^S are buy and sell volume per bucket (which is denoted by subscript b) and n represents number of buckets used for averaging. For exact derivation of VPIN, see Easley, Lopez de Prado and O'Hara (2012).

The calculation of VPIN depends on the length of the time bar used to calculate order imbalance and the number of buckets over which the value is averaged. Stating that there is en masse order misclassification in E-Mini futures using standard classification algorithms in a high frequency setting, Easley, Lopez de Prado and O'Hara (2012) suggest using bulk classification to determine buy and sell sides of trades in a given time bar. The bulk algorithm determines the trade imbalance by:

$$V_b^B = \sum_{i=j(b-1)+1}^{j(b)} V_i \cdot \phi\left(\frac{P_i - P_{i-1}}{\sigma_{\Delta P}}\right), \quad (3.4)$$

$$V_b^S = 1 - V_b^B, \quad (2.5)$$

where $j(b)$ represents the index for the last time bar in the b^{th} volume bucket, P is the price of asset, ϕ is the CDF of standard normal distribution and $\sigma_{\Delta P}$ is an estimate of return volatility.

This bulk procedure depends on the normal distribution of asset returns in physical time, which may not hold under the extreme conditions of a flash crash where trades are expected to be one-sided. The aggregation and averaging of the VPIN measures over an interval also might cause this toxicity measure to lag behind the real-time market dynamics of a flash crash in addition to introducing serial correlation to the series.

MH utilizes order book and trade variables found to be significant in Chapter 2, namely volume imbalance, duration and bid or ask depth. By combining these variables with the implications of the information-based market microstructure theory, the issue of forecasting liquidity based mini crashes will be addressed. However, the structure of the model is flexible enough to accommodate its use for "flash dashes" not so infrequently observed for single stocks (Golub, Keane and Poon (2012)).

MH takes into account the joint effects of volume, liquidity, order imbalance and duration on the evolution of financial time series. The proposed MH equation can also be linked to signed probability of informed trading presented in Equation (3.3) by the following equation:

$$MH = \left(\frac{Vol\ Imb}{Volume}\ x\ \frac{Volume}{Liquidity}\right)^{1/Duration} = \left(\frac{Vol\ Imb}{Liquidity}\right)^{1/Duration} \qquad (3.6)$$

where $Vol\ Imb$ is the volume difference between bid and ask initiated trades expressed in number of shares $(V^B - V^S)$, $Volume$ is cumulative trade volume, $Liquidity$ is the number of units standing on the bid side of the book waiting to be traded and finally $Duration$ represents average tick or order book update duration between two sampling points.

The above setup produces a simple way of interpreting MH in a high frequency setting. The total volume of trades ceases to be explicitly included in the equation. Instead it is the total available liquidity against the order imbalance between each k minutes that determines the flash crash probability. This is a much more intuitive gauge of impending sudden moves as price cascades are often caused by insufficient liquidity or open interest against a sustained order imbalance. As MH focuses on high frequency crashes, bid depth was selected to represent the liquidity conditions since the standing orders on the buy side of the order book would give an indication of the market participants' willingness to defend the asset price against sudden sell orders. Furthermore, unlike VPIN, MH captures time related information by explicitly making use of calendar time rather than using volume buckets. The inclusion of the duration term ensures MH accounts for the absorption limits of the market in physical time.

A total of 11 variables are formed using the trade and order book data. Ask depth and bid depth refers to the total volume standing on the order book at any given point. Order book imbalance refers to the difference between bid and ask depth. The incline variable is constructed using the cumulative volume on the order book in a step-wise fashion to estimate the slope of a best-fit line. For a detailed explanation of the construction of the incline variable, the reader may refer to Deuskar & Johnson (2011). The mean value for these order book variables are used for 5 minute estimates. Additionally, a 5-minute mean spread value is calculated as well.

In addition to volume, a volume imbalance term is formed by taking the volume difference between bid and ask initiated trades. Similarly, initiator imbalance is computed by taking the difference between the number of aggressors on buy and sell sides. The total number of trades is also computed as a separate measure.

The order book is updated each time a new order arrives, a standing order is altered or deleted or when a trade is realized. To reflect the dynamic nature of the order book, two duration parameters are created. "Duration" refers to the average order book update duration, whereas "trade duration" refers to the average time between realized trades in the 5 minute window.

The return series computed over 5-minute intervals are then transformed into one of the "crash" or "no crash" categories for a number of different crash thresholds.

Three different alternatives for predicting future crashes are used to assess the explanatory power of the proposed MH equation. The first alternative entails a simple linear combination of the variables formed via trade or order book information. To be specific, the linear estimator is of the form:

$$LinF = \beta_1 \left(1/_{Duration}\right) + \beta_2 \, Vol\ Imb + \beta_3 \, Liquidity \tag{3.7}$$

A measure similar to the VPIN measure outlined in Easley, Lopez de Prado and O'Hara (2012) is used as the second alternative. However, there are subtle changes to the construction of this measure compared to the original VPIN. Instead of employing the widely used Lee & Ready (1991) algorithm or the bulk classification described in Easley, Lopez de Prado and O'Hara (2012) to classify trades, trades are classified via the tags provided by CME. The use of 1-min time bars and volume bucketing are also rendered redundant by this approach since VPIN computed in this study uses time buckets to make a fair comparison with other metrics. These changes essentially remove artificially introduced errors and help construct a comparable VPIN measure. Finally, the third method that is put to test is the MH equation.

The ability of each of the three measures described above is put to test using a linear discriminant analysis (LDA). The general form of the discriminant equation used in LDA is:

$$D = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_{n-1} X_{n-1} + \beta_n X_n \tag{3.8}$$

where D is discriminant value, X_i represent variable vectors and β_i are the associated weights of each corresponding variable and β_0 is a constant.

LDA maximizes the following objective function:

$$J(\mathbb{D}) = \frac{\mathbb{D}^{\intercal}\mathcal{M}_b\mathbb{D}}{\mathbb{D}^{\intercal}\mathcal{M}_w\mathbb{D}} \tag{3.9}$$

where $\mathbb{D}$ is the direction that maximizes class separability and M_b and M_w represent between-class and within-class scatter matrices respectively (Fisher (1936)). The scatter matrices can be expressed explicitly as:

$$\mathcal{M}_b = \sum_j (\mu_c - \tilde{\mu})(\mu_c - \tilde{\mu})^{\intercal} \tag{3.10}$$

$$\mathcal{M}_w = \sum_i (x_i - \mu_c)(x_i - \mu_c)^{\intercal} \tag{3.11}$$

where μ_c represents class mean, $\tilde{\mu}$ represents pooled mean and x_i represents individual data points for a given class.

Put in simpler terms, LDA is a dimensionality reduction technique. It aims to classify individual data points into classes by projecting them on a scalar. Thus, when the objective function in Equation (3.9) is maximized, what one essentially does is to find a projection vector that will put observations within the same class close together while keeping class means as distant from each other as possible.

LDA functions for the linear, VPIN and MH methods respectively can then be expressed as:

$$D_{Linear} = \beta_0 + \beta_1 \, {}^{1}/_{dur} + \beta_2 \, Vol\ Imb + \beta_3 \, liq \tag{3.12}$$

$$D_{VPIN} = \beta_0 + \beta_1 \frac{|Vol\ Imb|}{vol} \tag{3.13}$$

$$D_{MH} = \begin{cases} \beta_0 + \beta_1 \left(\frac{Vol\ Imb}{liq}\right)^{1/dur}, & Vol\ Imb \geq 0 \\ \beta_0 + \beta_1 \left[-1 \cdot \left(\frac{Vol\ Imb}{liq}\right)^{1/dur}\right], & Vol\ Imb < 0 \end{cases} . \tag{3.14}$$

In each of the three cases, upon finding the weights, trades are classified into one of the two possible outcomes using LDA and compared with the actual results. The binary classification produces the following confusion matrix:

Table 3.1: Confusion Matrix

	Crisis	**No Crisis**
Signal	True Positive **TP**	False Positive **FP**
No Signal	False Negative **FN**	True Negative **TN**

An ideal classifier with perfect foresight would only produce results along the TP-TN diagonal. To assess the success of each methodology, “classification accuracy” is chosen as the primary performance measure. Classification accuracy (CA), which is the ratio of correctly specified observations to total number of observations, can be expressed as:

$$CA = \frac{(TP+TN)}{(TP+FP+FN+TN)} . \tag{3.15}$$

Precision (PR) and recall (RE) are used as additional measures of binary classification. The two performance measures can be expressed as:

$$PR = \frac{TP}{(TP+FP)} , \tag{3.16}$$

$$RE = \frac{TP}{(TP+FN)} . \tag{3.17}$$

Finally, an in-sample linear discriminant analysis, using a range of crash thresholds, is conducted to test the robustness of results. To further examine the performance of the proposed methods, an out-of-sample LDA is performed using a rolling window approach.

3.4 Data

Two different high frequency datasets are used in this chapter. The first dataset on the Flash Crash is provided by the Chicago Mercantile Exchange (CME). CME uses a continuous-time double auction system where the e-Mini future contracts on the S&P500 stocks are traded. At any given point in time during the trading day, the Level 2 data provided by CME will entail orders 10-deep in the order book with information of volume and price of standing orders ranked first by price competitiveness and then by time of submission. Additionally, realized trades with volume and time information are also available in the same dataset. This data is used to assess both in-sample and out-of-sample accuracy of MH. The in-sample CME analysis takes into account two full trading weeks surrounding the Flash Crash, specifically the sample runs from May 3rd to 14th of May 2010. The out-of-sample CME analysis uses these 2 weeks as the training set. To keep the total number of observations used in estimations constant in the out-of-sample approach, the oldest observation is dropped each time a new one is added. Thus, the trailing window approach used in the out-of-sample analysis forms the forecast vector sequentially with single period ahead forecasts and runs to 30th of May 2010. The 31st of May 2010 is not included in the dataset due to the Memorial Day.

For both in-sample and out-of-sample analysis only data points that were realized during S&P500 trading hours, 9:30 – 16:00, were included in this research. Since the futures market is open almost around the clock, this truncation was necessary as trading volume falls dramatically after market close. Since liquidity is a key component of MH, using data from illiquid off-market hours would impair inferences drawn from this analysis. Hence, only normal trading hours data is used in this research.

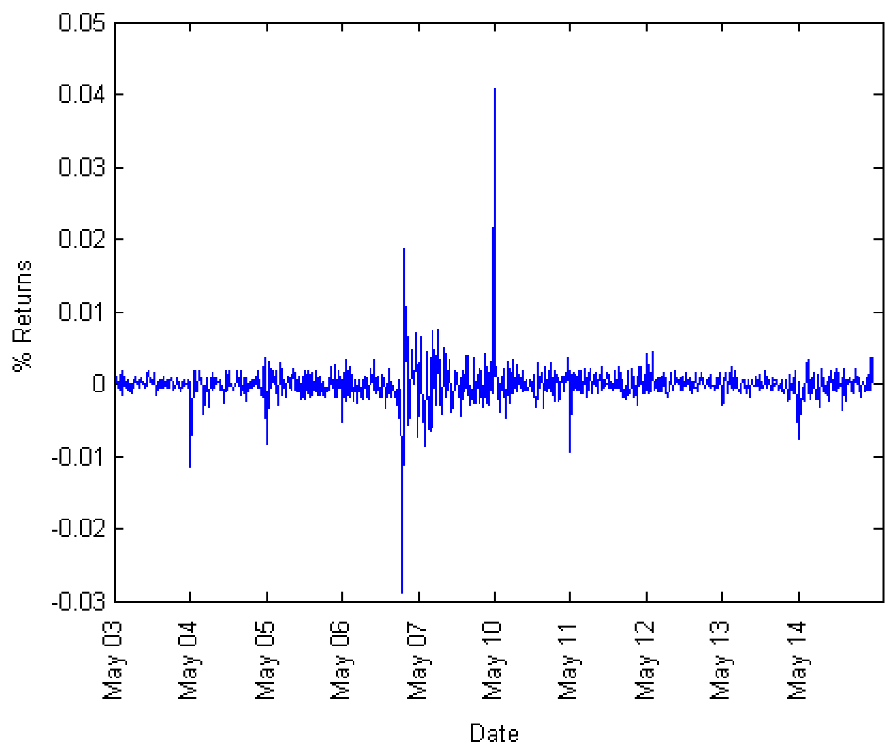
The second high frequency dataset includes SETS data provided from the London Stock Exchange (LSE) where single stocks are traded in a similar fashion to CME. The Level 2 dataset includes the whole order book depth at any given point in time which may vary considerably with the time of the day. The order book data includes only "public" orders that appear on the order book and excludes order types such as non-persistent or iceberg orders as well as OTC trades. Similar to CME, only the trading hours for LSE, 08:00 – 16:30, are used for this analysis. Four liquid stocks, namely HSBC, BG Group, British Petroleum and British American Tobacco are selected. To make a fair comparison with CME results and not to include irrelevant data in the estimation procedures, a 2-week moving window is selected as the training period to form the flash crash estimates for the rest of the month. Out-of-sample analysis for each stock is conducted separately on a monthly basis using a range of crash thresholds to test for robustness. The time period for LSE stocks runs from July 2007 to June 2008.

3.5 Results

The Flash Crash is characterized by a rapid plunge of e-Mini futures which was followed by a similar trend in the spot market and a recovery following the trigger of circuit breakers. The S&P500 future prices had lost approximately 3% of its value in less than 5 minutes. This fact is clearly observable in the five minute returns plot depicted in Figure 3.1.

Figure 3.1: e-Mini S&P 500 Futures Returns Sampled Every 5-Minutes

(May 3rd to May 14th 2010)



As a preliminary check for any explanatory power of the proposed variables, two separate regression analyses are conducted using contemporary and lagged variables. The results for this descriptive analysis are presented in Table 3.2 and Table 3.3 respectively.

Table 3.2: Simple Linear Regression of Contemporary Variables on e-Mini S&P500 Futures Returns - Sampled Every 5-Minutes (May 3rd to May 14th 2010)

variable	constant	coefficient	R^2
ask depth	-2.2009 e-4	1.3926 e-8	0.0009
bid depth	3.1267 e-4	-3.6566 e-8	0.0046
order book imbalance	-2.5475 e-4	-1.1168 e-7	0.0214
incline	1.7131 e-5	1.6195 e-6	0.0002
initiator imbalance	3.8417 e-5	4.5353 e-7	0.1717
inverse duration	2.3682 e-4	-4.9540 e-6	0.0042
trade duration	-2.1931 e-5	-6.4981 e-6	0.0002
number of trades	-3.8622 e-5	-1.0920 e-9	0.0017
spread	-0.0106	4.1985 e-4	0.0290
volume imbalance	-7.9440 e-6	8.6517 e-8	0.1955
volume	-3.7207 e-5	-2.8054 e-10	0.0017

Table 3.3: Simple Linear Regression of Lagged Variables on e-Mini S&P500 Futures Returns Sampled Every 5-Minutes (May 3rd to May 14th 2010)

variable	constant	coefficient	R^2
ask depth	2.6407 e-4	-2.7086 e-8	0.0035
bid depth	-7.8079 e-6	-4.8478 e-9	0.0001
order book imbalance	6.1611 e-5	6.6240 e-8	0.0075
incline	1.5308 e-4	4.7019 e-6	0.0014
initiator imbalance	-4.8079 e-5	4.1231 e-8	0.0014
inverse duration	2.4848 e-6	-1.0022 e-6	0.0002
trade duration	-2.8216 e-4	4.4254 e-5	0.0083
number of trades	-4.8849 e-5	-5.0929 e-10	0.0004
spread	-0.0205	8.1384 e-4	0.1089
volume imbalance	-5.2073 e-5	8.0172 e-9	0.0017
volume	-4.9398 e-5	-1.1203 e-10	0.0003

As expected, there is a considerable loss of explanatory power in most of the variables. Table 3.2 shows that volume imbalance and initiator imbalance terms, with respective R^2 values of 0.1717 and 0.1955, explain almost 20% of contemporary returns. When lagged, these variables lose most of their explanatory power and only achieve R^2 values of 0.0014 and 0.0017, respectively. Trade duration terms, on the other hand, now has a higher R^2 value, 0.0083 compared to the previous 0.0002. This finding further supports the case for using duration terms in the MH equation. Additionally, the bid-ask spread achieves very high R^2 values in both contemporary and lagged regressions, 0.0290 and 0.1089 respectively. However, the spread variable is not included in further analysis as it is non-stationary.

In the next stage, by using LDA, the three alternatives detailed in Equations (3.12) - (3.14) are compared with respect to their classification accuracy, precision and recall. Tables 3.4 - 3.7 represent CME in-sample and out-of-sample results.[20]

Table 3.4: Linear Crash Estimator LDA Results for E-Mini S&P500 Futures[21]

Crash	**Accuracy**		**Precision**		**Recall**	
Threshold	In-Sample	Out-Sample	In-Sample	Out-Sample	In-Sample	Out-Sample
-0.25%	78.7%	66.8%	16.8%	10.3%	63.8%	44.1%
-0.50%	89.2%	86.0%	9.7%	1.9%	100.0%	33.3%
-0.75%	88.4%	93.5%	3.2%	0.0%	100.0%	0.0%
-1.00%	90.0%	97.7%	2.5%	0.0%	100.0%	N/A

[20] The confusion matrices used to prepare Tables 3.6 – 3.9 are presented in Appendix D.

[21] N/A values for recall indicate that there were no crashes registered in the actual data using chosen crash threshold.

Table 3.5: VPIN LDA Results for E-Mini S&P500 Futures[22]

Crash Threshold	Accuracy		Precision		Recall	
	In-Sample	Out-Sample	In-Sample	Out-Sample	In-Sample	Out-Sample
-0.25%	55.6%	49.2%	9.0%	8.2%	70.2%	55.9%
-0.50%	52.0%	54.4%	1.6%	1.1%	66.7%	66.7%
-0.75%	54.2%	79.0%	0.6%	0.0%	66.7%	0.0%
-1.00%	48.8%	80.6%	0.5%	0.0%	100.0%	N/A

Table 3.6: Market Heat LDA Results for E-Mini S&P500 Futures

Crash Threshold	Accuracy		Precision		Recall	
	In-Sample	Out-Sample	In-Sample	Out-Sample	In-Sample	Out-Sample
-0.25%	93.1%	71.2%	31.6%	3.9%	12.8%	11.9%
-0.50%	98.3%	87.8%	25.0%	1.1%	22.2%	16.7%
-0.75%	97.4%	96.2%	5.3%	0.0%	33.3%	0.0%
-1.00%	98.2%	100.0%	7.1%	N/A	50.0%	N/A

Table 3.7: Market Heat LDA Results for E-Mini S&P500 Futures Using Trade Duration

Crash Threshold	Accuracy		Precision		Recall	
	In-Sample	Out-Sample	In-Sample	Out-Sample	In-Sample	Out-Sample
-0.25%	81.0%	61.9%	12.0%	6.3%	34.0%	28.8%
-0.50%	97.0%	84.0%	18.2%	1.6%	44.4%	33.3%
-0.75%	93.6%	96.0%	2.0%	0.0%	33.3%	0.0%
-1.00%	97.3%	99.5%	4.8%	0.0%	50.0%	N/A

The analysis for both in-sample and out-of-sample S&P500 futures point to a general increase in classification accuracy and a decrease in precision and recall for all methodologies as crash threshold is increased. Furthermore, in-sample CME results point to a striking conclusion. Linear, VPIN and MH all signaled for the Flash Crash ex-ante, using in-sample estimated β's and data 5 minutes prior to the Flash Crash. This predictive ability of all three models is attributable to the high recall values registered despite low

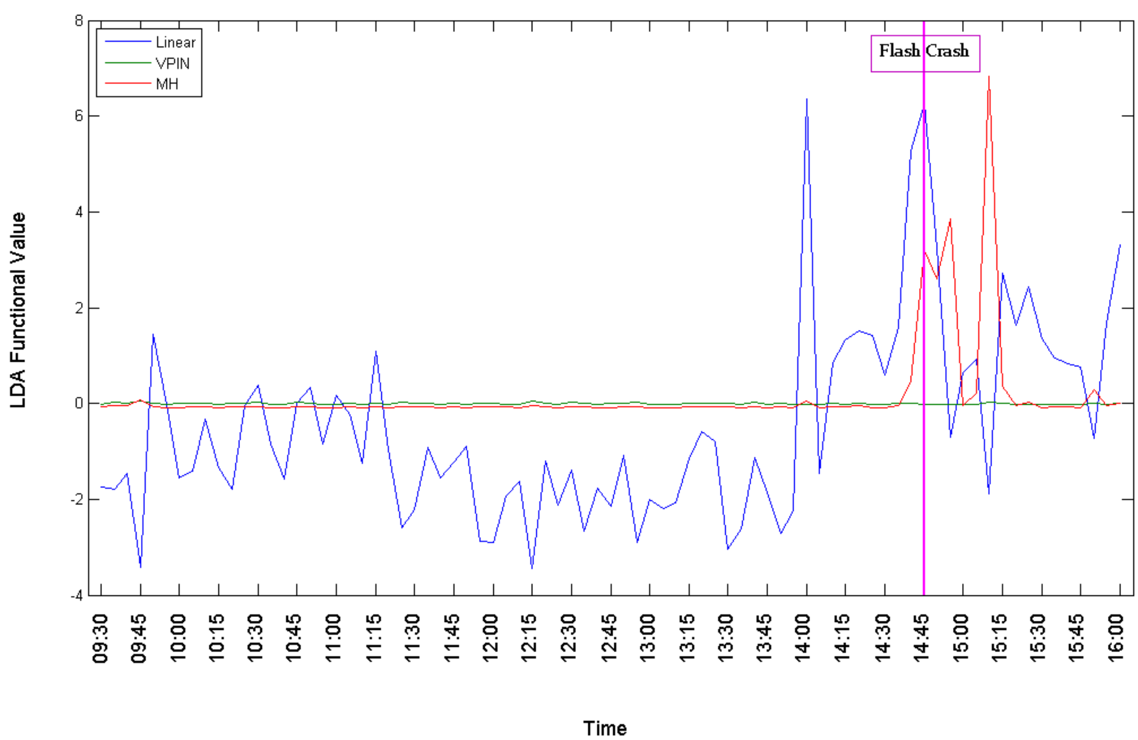
[22] The dramatic increase in VPIN's classification accuracy in the out-of-sample data is mostly due to the reduced number of crashes registered in the actual data.

precision. However, low precision values are to be expected in a heavily unbalanced dataset such as the one used here, where the binary classifier is not set to give more weight to false positives.

Although all three models were similar in predicting the Flash Crash, there were stark differences in the functional values obtained by each methodology. Although VPIN signals for the Flash Crash, the functional value attained by this measure at the time of the event, 14:45, is not radically higher than the average for the series. On the other hand, LDA values for linear and MH methods surge significantly during this timeframe. Figure 3.2 shows the LDA values obtained by the three approaches on the day of the Flash Crash.

Figure 3.2: LDA Values for 3 Alternatives on May 6th 2010

(Using Lagged Independent Variables)



Additionally, among the three approaches studied here, VPIN consistently produced the lowest classification accuracy with respect to different crash thresholds. This could essentially be due to two reasons. Table 3.2 and Table 3.3 show little explanatory power for volume of trades both in contemporary and lagged analysis. Hence, including it in a crisis measure may not bring any additional value to crash prediction. Furthermore, taking the absolute value of the volume imbalance term, which loses a considerable part of its explanatory power when lagged, might further dilute the information carried by this component. However, comparison between MH computed using signed vs. absolute value of volume imbalance term shows almost no loss of information. In fact, CME in-sample analysis represents the only occasion where there is a difference between the results for absolute and signed volume imbalance. Hence, only the results using absolute value of volume imbalance is reported.

Contrary to VPIN, MH produces the highest classification scores for all three crash definitions. Its advantages in CA become starker as lower thresholds are applied to the data. Moreover, the number of false positives produced by MH is dramatically lower compared to both the linear and VPIN models. While VPIN classified almost half the trades as "crashes", the number of crash predictions made by MH stood at a mere 2%.

A separate MH measure is also constructed using average trade duration instead of order book update duration to compare the effects of different duration gauges. The results for this new MH measure are presented in Table 3.7. The performance difference between the two MH equations is stark. Classification accuracy falls dramatically while the number of crash predictions increases multifold. This large difference in between two MH approaches shows that it is the average order book update duration that carries the bulk of the information needed to predict crashes. The patterns observed using in-sample data extend directly to the out-of-sample analysis. Naturally, all three methodologies have lower values for classification accuracy, precision and recall. However, this is an anticipated outcome given the use of an extended set of data points.

What is remarkable however, is the consistency of results across different markets. The findings in the US futures market also extend to all selected single stocks traded on the London Stock Exchange, though there is one methodological difference. In certain months there were instances where all trade volume was initiated by the bid or ask side of the order book. This did not constitute a problem for linear and MH methods, but on such occasions the discriminant analysis was unable to classify these data points for VPIN as by construction VPIN allows values to appear within the range of 0-1. These data points were removed for all three types of analysis in such cases. The amount of data removed in a given month ranged from 1-10% of the whole dataset, which could amount to a significant loss of information if one were to use VPIN.

Figures 3.3 – 3.5 show summary results for classification accuracy with respect to different crash thresholds in the out-of-sample analysis for the LSE stocks. Complete set of results for classification accuracy, precision and recall values for LSE stocks are presented in Appendices E - G. The out-of-sample performance for LSE stocks is very similar to that of E-Mini futures. Although the overall variability of results across months and different stocks is higher for single stocks compared to S&P500 futures data, VPIN continued to consistently yield the lowest CA scores. Similarly, MH continued to result in the highest classification accuracy most of the time although its benefits compared to the linear approach was not as stark as in the case of CME. At times the linear approach yielded better results, the difference between the linear and MH methods were slight. These incidents where the linear method produced the better results could be due to the idiosyncratic nature of single stocks or perhaps it could be directly caused by lower market activity in single stocks compared to S&P500 futures. Thus, using 5-minute sampling interval for single stocks may have resulted in loss of information. However, the effect of sparse sampling on single stock crash identification is beyond the scope of this study.

Figure 3.3: Classification Accuracy for LSE Stocks (Crash Threshold: -0.25%)

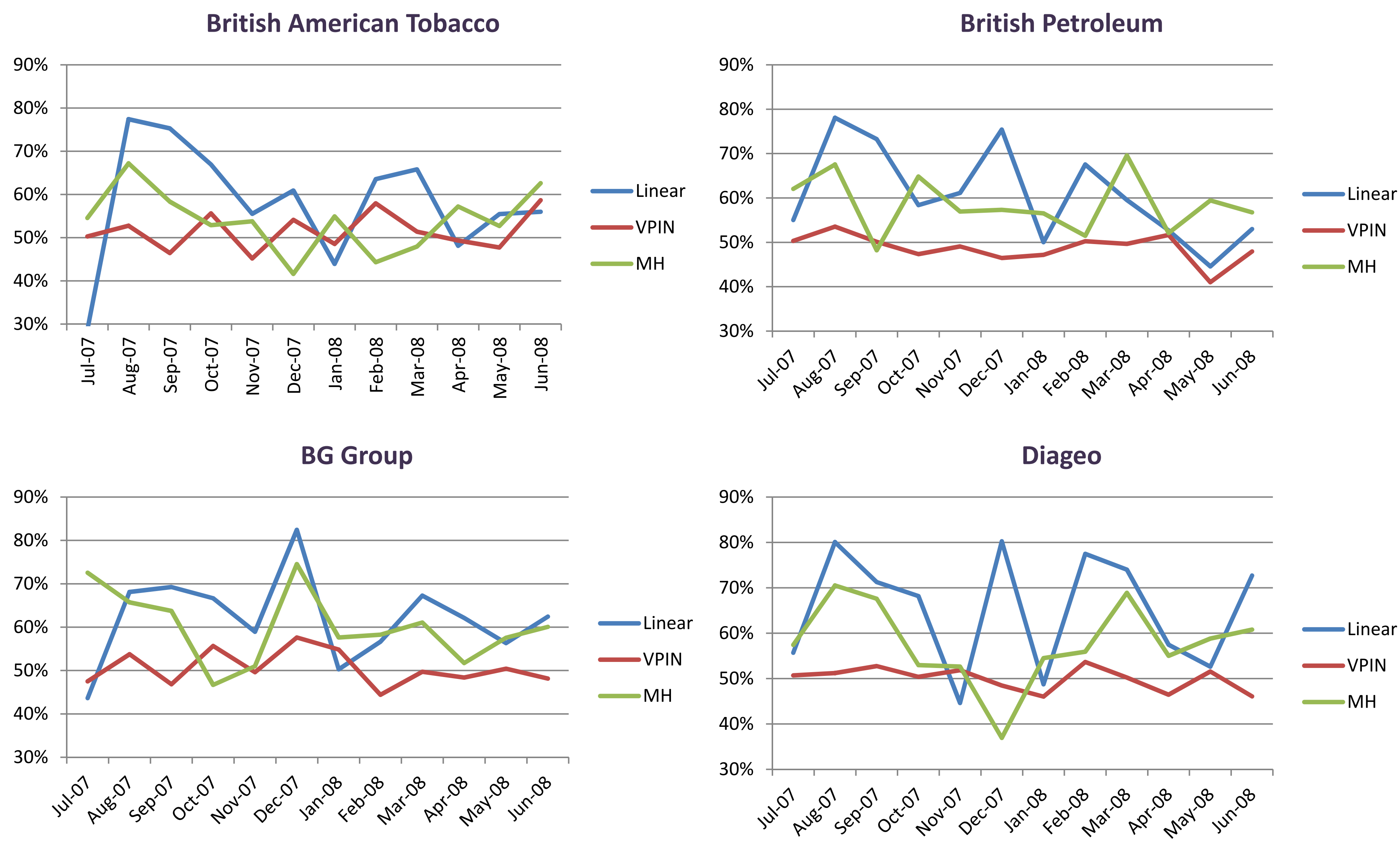


Figure 3.4: Classification Accuracy for LSE Stocks (Crash Threshold: -0.25%)

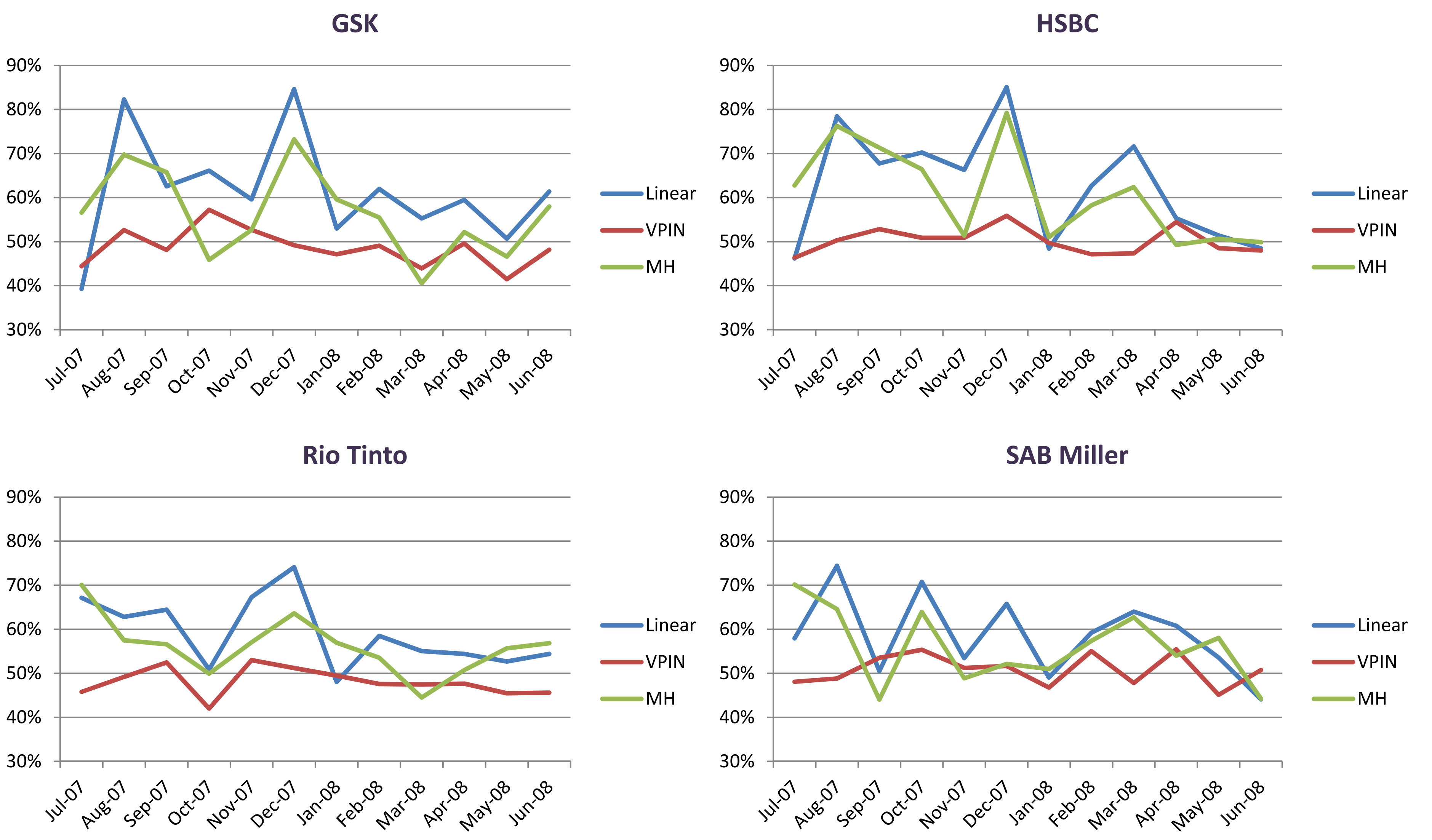
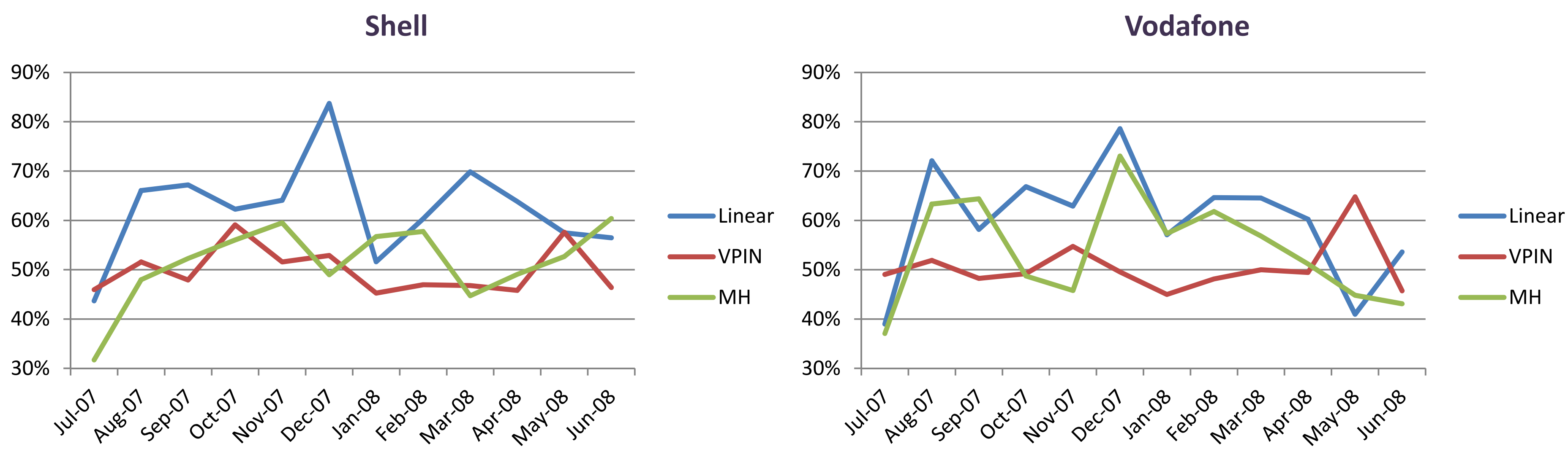


Figure 3.5: Classification Accuracy for LSE Stocks (Crash Threshold: -0.25%)



3.6 Discussion of Results

This chapter focuses on sudden price moves in two technically and geographically different financial markets, namely E-Mini S&P500 futures and selected LSE stocks. Despite numerous differences in market structures, the findings regarding the predictability of "flash crashes" extend to both markets under consideration.

The preliminary analysis shows variables used in the construction of VPIN lose considerable explanatory power when lagged. But even then, prediction of large scale events such as the Flash Crash remains an undemanding task. The ability of a naive linear function in predicting the Flash Crash underscores a simple fact: the Flash Crash could have been avoided if the necessary warning systems were in place.

Comparison between MH using signed vs. absolute value of the volume imbalance term shows results are almost identical. This close similarity in outcomes points to the necessity of explicitly including a liquidity gauge in any crash metric as volume imbalance alone provides only part of the essential information. Additional analysis on a different version of MH, namely one that uses average trade duration instead of average order book update duration, yields an interesting result. The difference between these two MH methods shows that the bulk of the information on an imminent crash is being carried by the average order book update duration, hence the poor performance of this secondary method using trade duration.

The market heat measure presented here, which takes into account both order imbalances in the volume of trades and the frequency with which the order book is updated, produces vastly superior results compared to all its counterparts. The ability of

MH to consistently outperform VPIN persists across different thresholds and markets. In fact, VPIN is almost always the worst performer with respect to the three performance measures used here. The original VPIN measure used in Easley, Lopez de Prado and O'Hara (2012) introduces artificial errors with the use of time bars and volume buckets, a practice which was averted here with the use of trade initiator tags. Nonetheless, the use of VPIN still remains problematic, as it requires additional data cleaning before conducting the final discriminant analysis.

MH outperforms the linear approach as well. However, the out-of-sample performance of MH depends on the economic climate and the asset class. Hence, the differences in between MH and the linear approach using LSE single stock data are not as pronounced as they are when using CME data. This could possibly be due to the idiosyncratic nature of single stocks or the notable lack of confidence in financial markets during the sampling period for the single stocks.

In conclusion, MH offers a new way of predicting episodes as dramatic as the Flash Crash. By utilizing the findings of Chapter 2 while determining the variables influencing price formation, MH successfully predicted most of the short term financial crashes, a primary goal of this thesis. Hence, Chapter 3 contributes to the literature by not only introducing a flash crash prediction tool that is more accurate than all other methods reported in the literature but also by adding further support to the use of order book variables in determining short term price movements in high frequency data.

The results point to a robust measure that is capable of outperforming its alternatives across different markets and crash thresholds. MH could have prevented the Flash Crash if it were employed as a circuit breaker at the time and given its excellent out-of-sample performance, MH may also prevent similar future episodes if it were to be incorporated into stock market warning systems.

A shortcoming of MH however, is its low precision despite its high recall. Thus, stock markets that are concerned with too many false alarms may instead use MH not to halt trading but to start watching the markets for additional crash indicators. Alternatively, MH could also be used to warn the designated market makers to supply further liquidity.

In the next chapter, I extend the crash prediction techniques to the macroeconomic level. By creating early warning systems for currency crashes, Chapter 4 augments the short-term nature of the predictions made in Chapter 3 and completes the thesis objective of crash prediction with different time horizons.

Chapter 4: The Missing Link in Early Warning Systems

4.1 Introduction

Chapter 3 focused on flash crashes that can be explained by market microstructure effects and order book imbalances. However, the various forms of financial crises that we observe in today's financial markets are not primarily composed of these short-term price movements. In fact, much of the problems that economies face in the long run are caused by macroeconomic imbalances that manifest themselves as rapid currency devaluations. Hence, in order to create a complete gauge of financial crashes, Chapter 4 shifts the focus of crash prediction to the macroeconomic level by focusing on currency crashes. Specifically, this chapter aims to forecast these rapid currency devaluations in G10 countries by focusing on both macroeconomic imbalances and market liquidity conditions.

In this final chapter of the thesis, prediction of crises that surpass flash crashes by both the magnitude of losses and the period of recovery will be the primary objective. The occurrence of financial crises is not rare by all standards and the failure of developed economies to fully recover from the sovereign debt crisis that has engulfed Europe and the United States proves that the recent bid for financial liberalization and globalization must be taken with heed. The colossal destruction of value in the developed nations under adverse market conditions only hints at the possible turmoil for smaller open economies. Hence, it is ever more essential for investors to predict these turbulent periods in financial markets.

In order to predict currency crashes, several binary and panel estimation models are used in this chapter. A key contribution of this chapter lies in the addition of market variables, such as VIX and TED-Spread to the set explanatory variables used to create the models. The results provide a key insight: market variables are strong determinants of future currency devaluations. While almost all macroeconomic variables are found to be insignificant for binary models, market variables hold predictive power for both binary and panel estimations. Panel estimations formed with debt related macroeconomic variables and liquidity related market variables produce profitable currency strategies. Hence, Chapter 4 not only provides a comparison between alternative crash prediction models using new market variables, but also fulfills the second goal of this thesis by producing successful early warning systems for G10 currencies.

Despite the various manifestations of macroeconomic crisis, all forms share a common trigger; the sudden realization that the balance sheet of an entity, large enough to impact an economy, is in fact not balanced. Whether it is debt accumulation or a sudden dip in the value of the assets, the apparent inability of this financial entity to service its debts manifests itself as a financial crisis. In the end, the cost is borne by the "real" owner of the obligations, the government.

Given that the main cause behind financial crises is balance sheet problems, we can classify crises into two main categories namely; debt crisis and asset crisis. Sovereign and private entities alike may be subject to both types of crises. Although the characteristic causes for each type are different, they are not entirely independent of each other as it is often the case that one induces another.

For instance, excessive borrowing by governments may constitute a currency crash hazard in an open economy where the government sustains borrowing via a high interest rate regime which in turn reduces the economy's competitiveness and output (Reinhart, Reinhart and Rogoff (2012)). In this setting, expectations of a currency crash could create a self-fulfilling mechanism and cause a sovereign debt crisis which would instantaneously increase the government's foreign liabilities. However, the increase in the output levels would also cause future tax income to increase and the benefits for the economy may outweigh the cost of the crisis. On the other hand, if excessive foreign debt is accumulated by the private sector, the debt overhang may adversely affect the firms' (households') ability to invest (consume) further, cooling the economy (Lo and Rogoff (2014); Mian and Sufi (2014)). Similarly, hidden borrowing by the banks and financial intermediaries could result in a banking crisis and the slowing down of the economy via reduced available funding. The re-capitalization of banks by the government puts sovereign balance sheets under stress and may cause debt-crisis if sovereign debt stock is large enough (Reinhart and Rogoff (2011)).

As outlined by Krugman (2002; 2010), asset-side problems may also cause currency crashes. In this type of crisis, deleveraging due to an asset bubble may cause assets to substantially decrease in value which might bring private firms as well as banks to insolvency. Wide spread economic contraction and government guarantees on bank deposits may force the government to step in and provide emergency credit lines to one or both parties to ensure stability. What follows is an increase in government debt, the expansion of the money supply and reduced interest rates, all of which inevitably result in currency devaluation. The credibility of the government plays an important role in securing the necessary funds to contain such wide spread balance sheet induced crisis and the cost of this procedure can be excessive for smaller open economies. This is due to the inability of emerging markets to erode their debt stock via inflation as they heavily depend on foreign denominated debt.

Much like the bank runs of the early 20th century, the 2008 mortgage crisis is a grim reminder of economic contraction that will follow wide-spread asset-side problems in banks' balance sheets. Currency crashes that are caused by asset-side problems are the most destructive (Krugman (2010)), which justifies the use of fundamentals to predict economic fragilities and currency crashes. Additionally, the self-fulfilling nature of currency crashes must also be factored in via global market variables.

On the whole, despite the long list of divergent causes behind different types of crises, most often each is immediately followed by a strong depreciation of the domestic currency against its counterparts. This is the main motivation behind mapping of currency crashes which has occupied academics and policy makers alike for decades. Over the decades several promising empirical models have been developed to predict these large scale devaluations (Kaminsky, Lizondo and Reinhart (1998); Berg and Patillo (BP) (1999)). However, successful prediction of currency crashes still remains an elusive objective.

In this chapter, I diverge from the crash literature on several fronts. First, I take an indirect approach to contagion, the spread of a currency crash among countries of similar dynamics. The banking system, which commands the currency markets, essentially works as a liquidity provision mechanism. When financial institutions start to charge higher premiums over the no-risk sovereign alternative, it alerts market players to possible short-term balance sheet or debt problems within the banking system. This in return reflects the downturn in the global economy. Thus, any shock to the banking system resonates with the FX markets more than any macroeconomic indicator. Hence, TED-Spread and VIX, which by construction measure the stability of the banking system and market confidence respectively, capture crisis events. For this reason, VIX and TED-Spread are utilized as global variables which account for the changes in investor sentiment.

Second, unlike most early warning system (EWS) studies, which predominantly concentrate on long-term crash predictions for emerging markets, I focus on the currency movements for developed markets combining macroeconomic indicators with market variables. The 1-month forecast horizon used in this chapter also has the advantage of making market variables relevant. In fact, both VIX and TED-Spread are found to contain important information for binary and panel data estimations.

Third, I utilize several binary crash prediction methodologies as well as panel models to forecast the changes currency markets. A through sensitivity analysis accompanies each of the estimated models. The sensitivity of signals approach to changes in the number of signals required to indicate a crash is tested for different thresholds. In order to assess the effects of pooling, country-by-country predictions using logit and probit models are compared to their pooled counterparts, which treat all data as if it were coming from a single country. Each binary estimation is accompanied by a ROC curve, which provides a visual performance measure of the estimates for each FX pair. Adjusted ROC curves along with area under the curve (AUC) measures are present for panel estimations as well, which directly gauges the profitability of investment strategies.

The results indicate the need to increase the number of signals required for the signaling approach to perform well. Furthermore, logit and probit estimations demonstrate that single FX pair crisis predictions yield good results for only a few of the countries. I also include lagged crash indicators and lagged returns in binary and panel estimations to test for any dynamic behavior in currency returns. However, no evidence is found to support the use of lagged binary indicators. As such, dynamic binary estimators seem unwarranted. Panel estimations; on the other hand, show that by using market variables and a short forecast horizon of 1-month, one can generate excess financial returns.

Section 4.2 reviews key crisis prediction models that appear in the literature. In Section 4.3, the crash prediction methodologies are discussed. Section 4.4 describes the dataset and the results are presented in Section 4.5. A discussion of the results is presented in Section 4.6.

4.2 Literature Review

The currency crash literature has produced three generations of theoretical models. The first generation models are based on the speculative attack model of Salant and Henderson (1978) and assume perfect foresight where balance of payments crises occur deterministically. The governments run persistent fiscal deficits and rising debt concerns push investors to attack the domestic currency en masse. The heavily indebted government is then left to choose between depleting its foreign reserves in defense of the currency or forego the fixed exchange rate regime (Krugman (1979)). The second generation models ushered by Obstfeld (1986) add multiple equilibria to its predecessor within a self-fulfilling prophecy framework. Obstfeld (1996) further extends second generation models where the government minimizes a quadratic loss function of inflation and deviation from natural output level.

The inability of the first and second generation models to account for the 1997 Asian crisis brought about the third generation models which focus on the balance sheet risks of the financial sector. Corsetti, Pesenti and Roubini (1999) and Chang and Velasco (2000) examine the role of foreign debt and excessive borrowing in the banking sector. Excessive foreign bank borrowing can constitute a hidden form of sovereign external debt under blanket guarantees to banks which sustain government borrowing via domestic bond purchases. This type of debt crisis induces currency crashes which increase the foreign liabilities on the balance sheets of banks and adversely affect their ability to lend, prolonging the recession. This is in sharp contrast with the fast recovery period following currency devaluations predicted in the first and second generation models. Krugman

(1999) also focused on the effects of devaluation on private sector balance sheets and later extended the liability side balance sheet drawdown to the asset-side in Krugman (2002) where deleveraging of assets cause wide-spread insolvency and eventually an economic recession.

The first well-known attempt to model currency crashes was undertaken by Kaminsky, Lizondo and Reinhart (KLR) (1998). KLR developed a nonparametric signals approach in which a list of economic indicators that diverge from their "normal" levels prior to a crash were used to form an EWS. KLR's definition of a currency crisis includes both successful and unsuccessful speculative attacks and account for these components by utilizing an index of "exchange market pressure" which is a weighted average of monthly percentage exchange rate and foreign reserve changes. In the KLR setting, a crisis is said to occur when the exchange market pressure index crosses an arbitrary threshold value, in the KLR case 3 standard deviations. Then for a signaling horizon of 24 months, each indicator is set to issue a signal beyond a certain threshold whose value is determined by minimizing its signal-to-noise ratio. KLR found support for a number of indicators including foreign exchange reserves, real exchange rate, inflation, credit growth, trade balance and fiscal deficit.

KLR's indicators approach has found much support in the literature and several extensions to the original model were proposed to address its shortcomings. Change in the interest rates was included into the market pressure index by Hawkins and Klau (2000) whereas others changed the threshold value for the market pressure index (Aziz, Caramazza and Salgado (2000) and Edison (2000)). Edison (2000) also documented the inherent problem in this type of crisis definition, where sample dependence of standard deviation may “erase” past episodes of currency crashes. Alternative crisis definitions as a percentage of exchange rate depreciation has also been offered by Esquivel and Larrain (1998), Bruggemann and Linne (2002) and Kumar, Moorthy and Perraudin (2003).

A major shortcoming of KLR model was addressed by Berg and Patillo (BP) (1999). BP model extended KLR’s signals approach into a multivariate framework by using a composite index of indicators. This way, loss of information due to conversion of indicators into binary variables was avoided and assessment of individual indicator performance became possible. The composite index approach was used within a probit setting and a linear combination of indicators expressed as percentiles produced marginally better results compared to KLR approach.

Similar to the KLR model, BP model is much celebrated and extensions have been proposed. To account for the post crisis bias, exclusion windows have been proposed by Eichengreen and Rose (1996) and Demirgüç-Kunt and Detragiache (1998). However, removal of data during the recovery period results in loss of information and introduction of artificial serial correlation Abiad (2003). To remedy this shortcoming, multinomial logit and probit models have been proposed. Bussiere and Fratzscher (2006) and Ciarlone and Trebeschi (2005) employ a three state crisis definition with *tranquil*, *crisis* and *post-crisis* periods and find encouraging improvements in the forecast results. Nonetheless, the arbitrary determination of the exclusion window of the post-crisis period still remains a concern. Therefore, the gains in utilizing a multinomial model should be considered carefully since recoveries often do not take place as suddenly as currency crashes.

The pooling of various country specific data was later questioned by Berg et al. (2008). Bussiere and Fratzscher (2006) found mixed results for fixed vs. random effect logit models using findings of a core "groupable" country dataset with an all-inclusive set. Similarly, the effect of pooling on crisis thresholds in the KLR model was put to test by Davis and Karim (2008) and country specific thresholds were found to perform better in crisis prediction at the cost of higher Type II errors. On the other hand, pooled thresholds produced much reduced noise-to-signal ratios.

As an alternative solution to the transformation of indicators into binary signals in the KLR model, Peria (1999), Abiad (2003) and Bussiere and Fratzscher (2006) used Markov regime-switching models. Abiad found the overall performance of regime-switching models with non-constant probabilities to be similar to the BP model with Markov models estimating a higher percentage of the crisis periods. However, direct comparability is not possible since Abiad uses country specific time series data with a short out-of-sample period.

The failure of fundamental variables to predict currency crashes has been addressed by the use of extreme value theorem (EVT) as well. Cumperayot and Kouwenberg (2013) tested a large number of fundamental indicators and found that only the real interest rate was able to account for crises asymptotically. Nag and Mitra (1999) and Marghescu, Sarlin and Liu (2010) use of artificial neural networks (ANN) to predict currency crashes. The in-sample fit of ANN models are high, but results of Marghescu, Sarlin and Liu (2010) indicate that they only occasionally outperform the static probit models in predicting crises. ANN models are not without their drawback however. The number of hidden layers and neurons render them prone to overfitting and their “blackbox” nature makes identification of marginal effects of each indicator inscrutable.

As mentioned earlier, Kumar, Moorthy and Perraudin (2003) uses a different crisis definition, namely percentage exchange rate deprecation within a logit framework. This crisis definition is similar to a different strand of the literature that focuses on the link between carry trades and currency crashes. While examining the failure of uncovered interest rate parity (UIP) foe developed markets, Brunnermeier, Nagel and Pedersen (2008) document significant correlation between weekly carry trade positions and market variables, VIX[23] and TED-Spread[24]. Additionally, the authors find strong contemporaneous correlation between VIX and excess FX returns for quarterly forecasts. Jurek (2014) also finds that the crash neutral carry returns for dollar-neutral portfolios are statistically zero which implies that the options market account for the skewness risk in its entirety, a finding which supports further use of market variables such as VIX.

Kauppi and Saikonnen (2008) account for the autoregressive nature of currency crashes by including lagged binary and lagged index variables. Candelon, Dumitrescu and Hurlin (2014) extend the autoregressive approach to a dynamic logit setting using a rolling window procedure. They also address the country clustering problem with the use of

[23] VIX, short for Chicago Board Options Exchange Volatility Index, is a real-time volatility measure of S&P 500 stocks. Quoted in annualized percentage points, VIX is a weighted estimator of 1-month implied volatility using a range of index options.

[24] TED-Spread is a proxy for credit risk calculated as the interest rate difference between the 3-month US T-Bill and 3-month Eurodollar contract (LIBOR). Since T-Bills are dollar risk free, the TED-Spread measures the credit risk in the unsecured lending market. It can also be interpreted as commercial banks’ need for liquidity.

methods in Kapetanios (2003). In their model, Candelon, Dumitrescu and Hurlin (2014) find that dynamic specifications outperform both static logit and Markov regime-switching models within sample and though single period ahead forecasts are not satisfactory, multiple period predictions are outstanding for out-of-sample forecasts. The autoregressive structure employed in above described models inherently account for the findings of Tudela (2004) where the probability of recovery increases as the crisis period prolongs.

Contagion, the spread of a currency crash to neighboring countries in a given region, presents a major drawback for EWS and impairs fundamentals' ability to predict currency crashes. Various types of contagion, namely regional, trade partner and common creditor contagion, has been found to hold explanatory power to account for market participants' varying reactions to fragilities in different countries, Brüggemann and Linne (2002), Beckmann, Menkhoff and Sawischlewski (2006), Eichengreen et al. (1996), Reinhart et al. (2000) and Moreno and Trehen (2000). For this reason, a successful EWS needs to account for contagion by either direct inclusion of a regional contagion parameter or a global variable that will gauge the changes in investor sentiment.

Existing crash prediction models in the literature predominantly use long-term forecast horizons and focus on emerging markets which are already prone to currency crashes. Furthermore, most studies in the literature lack any short-term market variables. This is a significant shortcoming since market variables relate highly to the conditions in the market. A downturn in financial markets may reduce liquidity for market players and render currencies prone to crashes, supporting the case for multiple equilibria.

In order to address these shortcomings, I build several binary and panel models for 9 developed markets. A 1-month forecast horizon is used for each model combining macroeconomic and market variables. Furthermore, the data used in this chapter includes both a boom period and the global meltdown of 2008. All models are trained using the boom period which enables me to test whether or not the models introduced here would be able to predict the currency movements during the mortgage crisis. In addition to binary crash predictions, estimates for 1-month ahead FX returns are evaluated and found to produce excess returns.

4.3 Methodology

In this section, both binary and panel models are used to forecast currency crashes. EWS introduced in this section incorporate dynamic and autoregressive specifications as well using both fundamental and market variables. The ability of these models to predict currency crashes is tested via binary classification performance measures as well as ROC curves.

Both macroeconomic and market variables are used to predict the movements in the FX rates. Macroeconomic variables include the change in the interest rate premium for each country above the US interest rates, the change in inflation (CPI), the change in unemployment, the change in the current account/GDP, the change in the reserves/GDP, the change in the money supply M2/GDP and GDP growth. Several of these macroeconomic variables were found to hold explanatory power in studies such as KLR, BP, Kumar, Moorthy and Perraudin (2003) and Burnside, Eichenbaum and Rebelo (2008). Market variables include the change in VIX, TED-Spread and the main stock market index returns. Market variables, VIX and TED-Spread account for the "heat" of the market and are intended as global gauges of investor sentiment. Hence, they can be used to explain the elevated sensitivity of market players to balance sheet problems and capital flight, manifestations often described with contagion parameters.

Additionally, lagged binary and index variables as in Candelon, Dumitrescu and Hurlin (2014) are used to check for any persistence in currency movements. All explanatory variables are tested for stationarity and lagged one period to predict crashes.

To test how the chosen macroeconomic and market variables determine the changes in FX rates, I start with a multiple linear regression run on each FX pair separately. The multiple linear regression containing a constant and all explanatory variables can be expressed as:

$$r_{ij} = \alpha_j + \beta_j X_{ij} + \varepsilon_{ij} , \tag{4.1}$$

where α_j, r_{ij}, X_{ij}, β_{ij}, ε_{ij} represent the constant, the FX returns, the vector of explanatory variables, the coefficients vector, and the error term for the i^{th} observation of the j^{th} FX pair, respectively.

Ordinary least squares approach is used to estimate α_j and β_j values for the in-sample period. The estimated α_j and β_j values are then used to predict the FX returns in the out-of-sample data. Both the in-sample and the out-of-sample estimates are accompanied by their corresponding adjusted coefficients of determination, R^2_{adj}, indicating how well the estimates explain the variation in each FX pair.

4.3.1 Binary Models

Given the importance of both market and fundamental variables the following setup is appropriate. A single crash definition, a loss of 2% in the FX rate, is employed for all binary models since the inclusion of non-crisis countries in the dataset produces "phantom" crises using the KLR approach, Kindman (2010). The cut-off point for currency crashes is also warranted by the shorter forecast horizon of 1-month. The use of a short term forecast horizon also validates the use of market variables as their effects will be considerable.

The signaling approach used in this study is similar to the one used in KLR. However, several changes have been made to the basic model. As previously stated, the crash definition is kept the same for all competing models tested here. Hence, the KLR definition for a crash, which also includes the changes in reserves, does not apply here. Furthermore, the set of indicator variables used to predict future crashes are different including market variables.

The first step when using the signaling method is to identify the crash periods for the in-sample data. Then, each indicator is tested separately to find the country-specific quantile which minimizes the signal-to-noise ratio. The signal-to-noise ratios are calculated using the confusion matrix presented in Table 3.1 where Type I and Type II errors can be interpreted as false and missed alarms respectively.

KLR actually minimized an adjusted signal-to-noise ratio, $\frac{\frac{FP}{FP+TN}}{\frac{TP}{TP+FN}}$, which is the ratio of false alarms as a percentage of all no crisis instances to the correct signals as a percentage of all crisis instances. KLR's adjusted signal-to-noise ratio is used to determine the percentiles used in this chapter. The percentiles range used in the search algorithm goes from 0.70 to 0.95 (0.05 to 0.30) for positively (negatively) correlated explanatory variables.

Then, unlike KLR, a trailing window approach is used to adjust the levels at which each indicator will signal a crash. Although the quantile for each indicator stays the same, the trailing window approach adjusts "levels" for the major changes seen during crash periods. Then, each time any one of the indicators signal a crash, it is assumed that a crash in the next month will occur. Yet, a single signal may not be the best way to indicate a crash and may often produce false positives. For this reason, different thresholds for the number of signals required to indicate a crash has been tested. Specifically, thresholds of above 0, 1, 2 and 3 were used. The results have been reported in three different measures of success, namely classification accuracy, precision and recall.

In addition to the signals approach, two other binary models, namely logit and probit, have been used to test binary models' predictive capabilities. Both of these models have similar functional forms and can easily be applied to any binary classification study. Given a binary crash indicator, C, which takes a vector of variables X as inputs, can be expressed as:

$$C_i = \alpha + \beta X_i + \varepsilon_i \;, \tag{4.2}$$

where α is the constant and X_i, C_i, ε_i, represent the crash predictor, the vector of explanatory variables and the error term for the i^{th} observation. The binary predictor C_i would signal a crash above a predetermined threshold t such that:

$$C_i = \begin{cases} 1, & C_i > t \\ 0, & C_i \leq t \end{cases} . \tag{4.3}$$

The threshold can be assumed to be 0 since the constant α is present in the model. Then probability of having an event can be expressed as:

$$Prob(C = 1) = Prob(C > 0) \tag{4.4}$$

It is trivial to show that Equation (4.3) is equivalent to the cumulative distribution function $F(\beta X)$, for a symmetric distribution.

The probit and logit models differ in their selected cumulative distribution functions. While probit uses the standard normal cumulative distribution function, logit uses:

$$F(C) = \frac{e^C}{e^C+1} \ . \tag{4.5}$$

Unlike the signals approach, the logit and probit models do not employ a trailing window approach. In addition to classification accuracy, precision and recall, the results for logit and probit models include a receiver operator characteristic (ROC) curve.

The ROC curve is a graphical representation of the performance of a binary classifier. Specifically, it focuses on the ratio of the true positive rate (sensitivity) and the false positive rate (1-specificity). TP and FP rates are mapped on the vertical and horizontal axes, respectively. The ROC curve is then constructed piecewise by varying the probability threshold. Since both TP and FP rates can go only up to 100%, the maximum attainable area is a unit square for a ROC curve. An ideal classifier, that sorts all data points into true positives and true negatives, would produce an inverted "L-shaped" curve covering the whole unit square, resulting in an area under the curve of 1. A 45° line also companies the curve for comparison, designating a random classifier with an AUC of 0.5.

4.3.2 Panel Models

Although currency crash prediction is of importance, this is not the only way to predict the movements in exchange rates. Panel models not only produce point estimations but also allow for more comprehensive analysis of the data. For this reason, fixed-effects and random-effects models were used to predict the percentage change in the FX rates. The random effects model, which assumes explanatory variables are orthogonal to a country's characteristic crash risk, can be expressed as:

$$r_{ij} = \mu + \beta_j X_{ij} + RE_j + \varepsilon_{ij} \, , \tag{4.6}$$

where μ is mean return for whole sample, RE_j is the country-specific random effect and ε_{ij}, represents the country specific error term for the i^{th} observation.

The fixed effects model, on the other hand removes the orthogonality assumption. Hence, for a return series of:

$$r_{ij} = \gamma_j + \beta_j X_{ij} + \varepsilon_{ij} \, , \tag{4.7}$$

where γ_j is the latent time-invariant country characteristic crash risk, the fixed effects model estimates the demeaned returns:

$$\left(r_{ij} - \bar{r}_j\right) = (\gamma_j - \bar{\gamma}_j) + \beta_j (X_{ij} - \bar{X}_{ij}) + (\varepsilon_{ij} - \bar{\varepsilon}_{ij}) \, . \tag{4.8}$$

Given γ_j is time-invariant, the term $(\gamma_j - \bar{\gamma}_j)$ is eliminated and Equation (4.8) simplifies to:

$$\left(r_{ij} - \bar{r}_j\right) = \beta_j (X_{ij} - \bar{X}_{ij}) + (\varepsilon_{ij} - \bar{\varepsilon}_{ij}) \, . \tag{4.9}$$

Since the panel methods described above do not produce any binary classifiers, a one-to-one comparison with binary models is not possible. However, the return weighted ROC curve introduced by Jordà and Taylor (2012) within their regime-switching vector error correction model presents a visual compromise. The adjusted ROC curve is constructed by taking into account the maximum gain (loss) one would make if they were to predict the direction of all returns to be positive (negative). By doing so, the attainable AUC for the adjusted ROC curve is reduced to 1, which allows for comparisons between similar models can be made.

4.4 Data

Chapter 4 focuses on currency crisis on a macroeconomic level which warrants the use of a different dataset that includes foreign exchange rates and macroeconomic variables at variable frequencies such as monthly, quarterly and yearly. The main focus of EWS literature has been emerging markets, which are prone to substantial foreign exchange fluctuations. However, the mortgage crisis of 2008 showed that wild FX swings are not exclusive of the G10. For this reason, 9 developed markets, specifically, United Kingdom, European Union, Switzerland, Sweden, Norway, Canada, Australia, Japan, and New Zealand were selected for analysis.

The United States constitutes a large part of all international transactions; as such the US dollar was selected as the basis for all exchange rates. Hence, all exchange rates are taken against the U.S. dollar and FX rates which are customarily quoted with US dollar in the numerator are inverted to make them consistent with the rest of the data. All macroeconomic and market data are sampled on a monthly basis. Macroeconomic variables which are released less frequently are adjusted by using simple linear interpolation. This method is warranted by frequently updated market surveys for the macroeconomic variables used here.

The dataset runs from January 2000 to December 2012 with a monthly sampling frequency. This period includes two distinct episodes during which the financial markets experienced both a boom and a global meltdown. The in-sample period is selected to include the boom period, which runs from January 2000 to June 2006. Consequently, the hold-out period runs from July 2006 to December 2012. This partitioning of the dataset gives one the opportunity to form an EWS which would truly signal an impending crisis, using information before any crash was on the horizon.

4.5 Results

Regression analysis on the dataset shows an expected result: there are significant differences between variables in their ability to explain FX returns across countries. Most macroeconomic variables do not have any explanatory power with respect to FX returns. Even then, there is no consistency in the macroeconomic variables that are found to be significant as they change significantly from country to country. Market variables on the other hand, present a different case. TED-Spread is often found to be a significant explanatory variable which is sometimes accompanied by VIX as well. Thus, it can be said that a simple linear regression of FX returns supports the case for using market variables in an EWS. I further test this assumption with additional models. The results for the multiple linear regressions for the whole dataset are presented in Tables I.1 – I.3 in Appendix I.

4.5.1 Binary Models

Binary crash prediction when compared to forecasting FX movements is a simpler undertaking. Nonetheless, the number of financial crashes and the wealth lost during these episodes show that we have not been able to create an EWS to avoid these crises. The following subsections represent the results for the binary models used in this chapter.

4.5.1.1 Indicator Approach

The indicator approach or the signaling approach is a straightforward method which allows me to test each variable's predictive capabilities. Tables 4.1–4.4 represent the indicator approach binary classification performance measures for varying number of the signals required to indicate a crash. Detailed results for the signals approach using out-of-sample data are presented in Appendix J.

Table 4.1: Indicator Approach with 1 or more signals

Country	Classification Accuracy	Precision	Recall
United Kingdom	0.4286	0.2453	0.7647
Eurozone	0.3636	0.2333	0.8235
Switzerland	0.3377	0.2647	0.9474
Australia	0.2468	0.1857	0.9286
Canada	0.3636	0.1818	0.7143
Japan	0.4805	0.2593	1.0000
Sweden	0.4156	0.2545	0.7778
Norway	0.4026	0.3016	0.9048
New Zealand	0.2727	0.2031	0.7222

Table 4.2: Indicator Approach with 2 or more signals

Country	Classification Accuracy	Precision	Recall
United Kingdom	0.6494	0.3333	0.5882
Eurozone	0.5195	0.2619	0.6471
Switzerland	0.5325	0.2927	0.6316
Australia	0.4286	0.2000	0.7143
Canada	0.6623	0.2857	0.5714
Japan	0.5974	0.2424	0.5714
Sweden	0.5195	0.2286	0.4444
Norway	0.4675	0.2917	0.6667
New Zealand	0.4545	0.2391	0.6111

Table 4.3: Indicator Approach with 3 or more signals

Country	Classification Accuracy	Precision	Recall
United Kingdom	0.7662	0.4762	0.5882
Eurozone	0.6364	0.3103	0.5294
Switzerland	0.6364	0.3448	0.5263
Australia	0.6104	0.2778	0.7143
Canada	0.8052	0.4667	0.5000
Japan	0.7532	0.3684	0.5000
Sweden	0.6623	0.2278	0.2278
Norway	0.6364	0.3478	0.3810
New Zealand	0.6623	0.3333	0.4444

Table 4.4: Indicator Approach with 4 or more signals

Country	Classification Accuracy	Precision	Recall
United Kingdom	0.7922	0.5385	0.4118
Eurozone	0.7143	0.3529	0.3529
Switzerland	0.7143	0.3846	0.2632
Australia	0.7273	0.2667	0.2857
Canada	0.8312	0.5714	0.2857
Japan	0.7792	0.4118	0.5000
Sweden	0.7143	0.3333	0.2222
Norway	0.7532	0.5714	0.3810
New Zealand	0.7403	0.4375	0.3889

Tables 4.1 - 4.4 show a clear pattern. As the number of signals required increases, classification accuracy and precision increase at the cost of recall. Table 4.1 shows persistent high rates of recall for all countries at 1 or more signals. Given the improvement in classification accuracy as the number of signals required to indicate a crash is increased, the use of a higher signal threshold seems warranted. Table 4.3 shows that the signaling approach was quite successful at predicting currency crashes with several countries attaining classification accuracy values above 60% when a minimum of 3 signals is required for a crash prediction.

4.5.1.2 Logit & Probit

The indicator approach has a limited ability to make 1-month ahead forecasts for currency crashes since the interactions between the indicators are not taken into account directly. This is a direct amalgamation of all the explanatory variables and gives a solid idea about which of the variables actually contribute to the model. Other binary models, such as the logit and probit, provide means to test the combined effect of the variables. The results for the in-sample logit and probit estimations for each country are shown in Appendix K where as Appendix L contains the results for in-sample pooled logit and probit estimations. The in-sample estimates for the coefficients are then combined with the explanatory variables observed during the out-of-sample period to arrive at the binary predictions. Tables 4.5 – 4.6 show the resulting out-of-sample binary performance measures for logit and probit regressions. Detailed results for out-of-sample country-by-country logit and probit estimations are presented in Appendices M and N, respectively.

Table 4.5: Country by Country Logit Results[25]

Country	Classification Accuracy	Precision	Recall
United Kingdom	0.7821	0.5000	0.1765
Eurozone	0.7468	0.3333	0.1765
Switzerland	0.7595	-	0.0000
Australia	0.6962	0.2727	0.4286
Canada	0.8228	-	0.0000
Japan	0.7975	0.0000	0.0000
Sweden	0.5949	0.2308	0.3333
Norway	0.7342	0.5000	0.1905
New Zealand	0.7692	-	0.0000

Table 4.6: Country by Country Probit Results

Country	Classification Accuracy	Precision	Recall
United Kingdom	0.7821	0.5000	0.3529
Eurozone	0.7468	0.3333	0.1765
Switzerland	0.7595	-	0.0000
Australia	0.6962	0.2727	0.4286
Canada	0.8228	-	0.0000
Japan	0.7975	0.0000	0.0000
Sweden	0.5949	0.2308	0.3333
Norway	0.7342	0.5000	0.1905
New Zealand	0.7692	-	0.0000

[25] When the model makes 0 crisis predictions, a non-applicable value (-) for precision and a zero (0.0000) for recall is observed. This is due to precision measuring the ratio of correct crisis predictions to the number of total predictions and recall measuring the ratio of correct crisis predictions to the number of actual crisis. Equations (3.16) and (3.17) show the respective formulas for precision and recall.

Appendices M – N show that the explanatory variables found to be significant for both logit and probit models are identical. However, Switzerland, Canada, Japan and New Zealand are found to be affected by none of the explanatory variables tested. As a result, no crisis predictions were made by these models. Consequently, precision and recall measures for these countries are 0 for both logit and probit models. This less than satisfactory performance of both logit and probit models suggests that these models may not be suitable for currency crash prediction despite the high classification accuracy values they obtain.

Similar to the indicators approach with 3 or more signals, logit and probit models forecast currency crashes with relatively high predictive power, where both binary approaches produce identical results. However, despite their accuracy, these two binary models do not use most of the macroeconomic variables. In fact, if we consider pooled logit and probit estimations, all macroeconomic variables become irrelevant and the change in VIX and TED-Spread are found to be the only two significant explanatory variables. Table 4.7 shows the performance measures for pooled logit and probit models. Detailed results for the pooled binary estimations using out-of-sample data are presented in Appendix O.

Table 4.7: Pooled Logit & Probit Results[26]

Model	Classification Accuracy	Precision	Recall
Logit	0.7692	0.4545	0.2941
Probit	0.7692	0.4545	0.2941

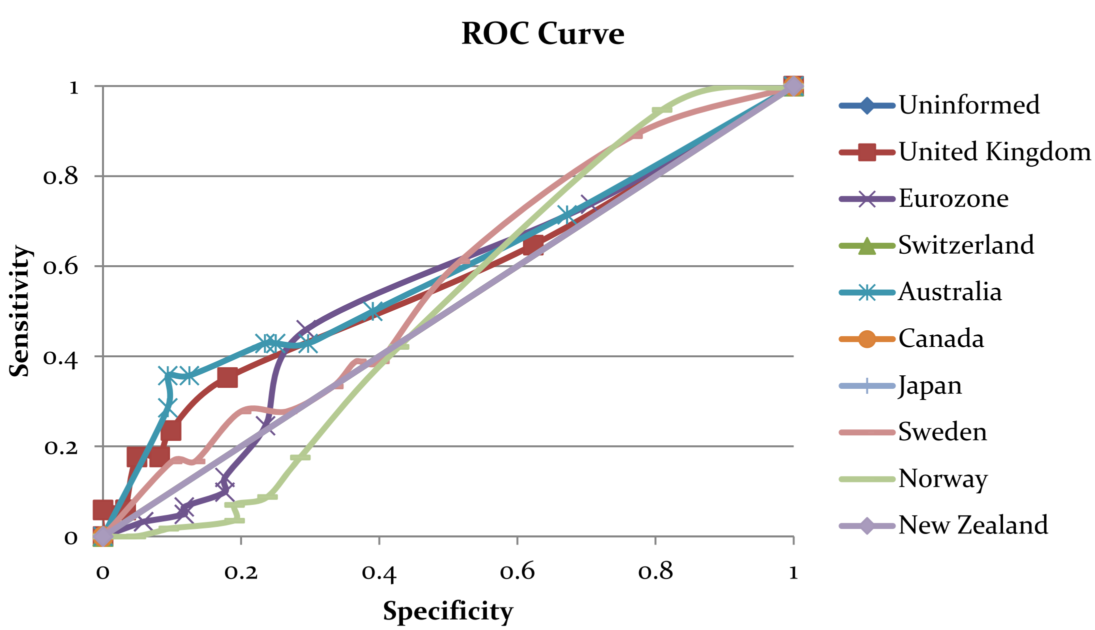
[26] The results for pooled logit and probit models are identical.

Given the total lack of macroeconomic variables in pooled logit and probit models and infrequent inclusion of macroeconomic variables in country-by-country logit and probit estimations, it can be argued that large monthly movements in currency markets may be rather short-sighted. Such myopic market behavior would cause liquidity concerns and immediate financial stability of the banking system to be direct contributors to currency crashes subduing the effects of a country's long term macroeconomic health.

In addition to the binary performance measures, the ROC curves for country-by-country and pooled logit and probit out-of-sample estimations are presented in Figure 4.1 – 4.3.

Figure 4.1: ROC Curves for Country-by-Country Logit Model[27]

(Out-of-Sample Estimation)



[27] The "Uninformed" variable in all the ROC curves contained in this thesis represents a random classifier with an AUC of 0.5.

Figure 4.2: ROC Curves for Country-by-Country Probit Model

(Out-of-Sample Estimation)

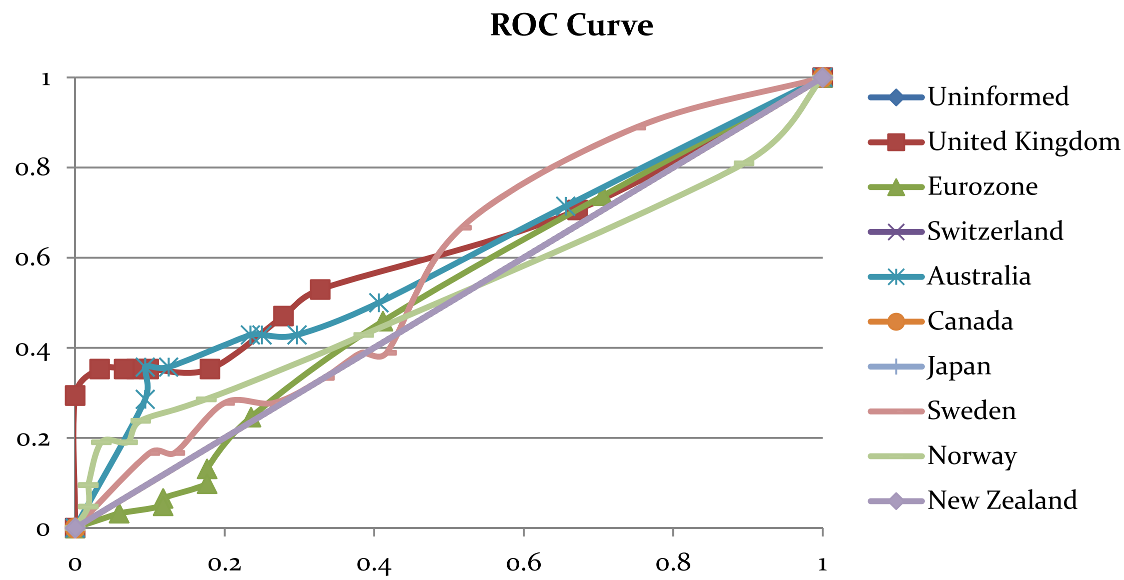
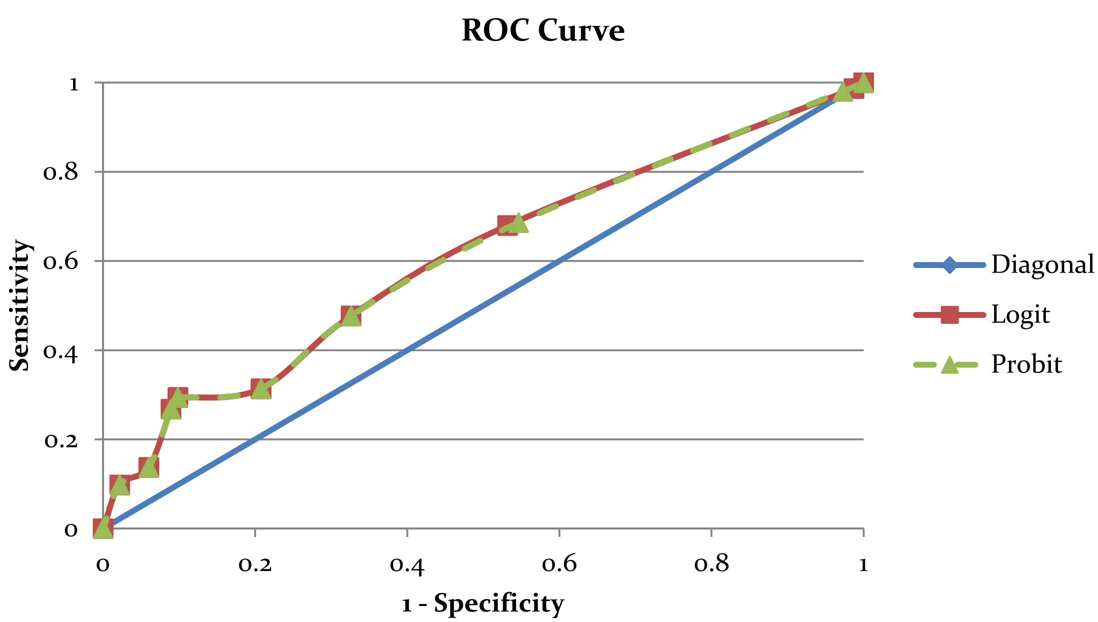


Figure 4.3: ROC Curves for Pooled Logit and Probit Models

(Out-of-Sample Estimation)



Figures 4.1 and 4.2 show that estimations for the United Kingdom and Australia outperform a random classifier reasonably well, while the estimations for the rest of the currency pairs are not satisfactory. Furthermore, compared to their country-by-country counterparts, pooled binary estimations, which only use market variables, result in well-behaved ROC curves. The areas under the curves for logit and probit estimations shown in Figure 4.3 stand at 0.6053 and 0.6045, respectively. These results are clearly superior to 0.5, the expected AUC for an uninformed crash identifier.

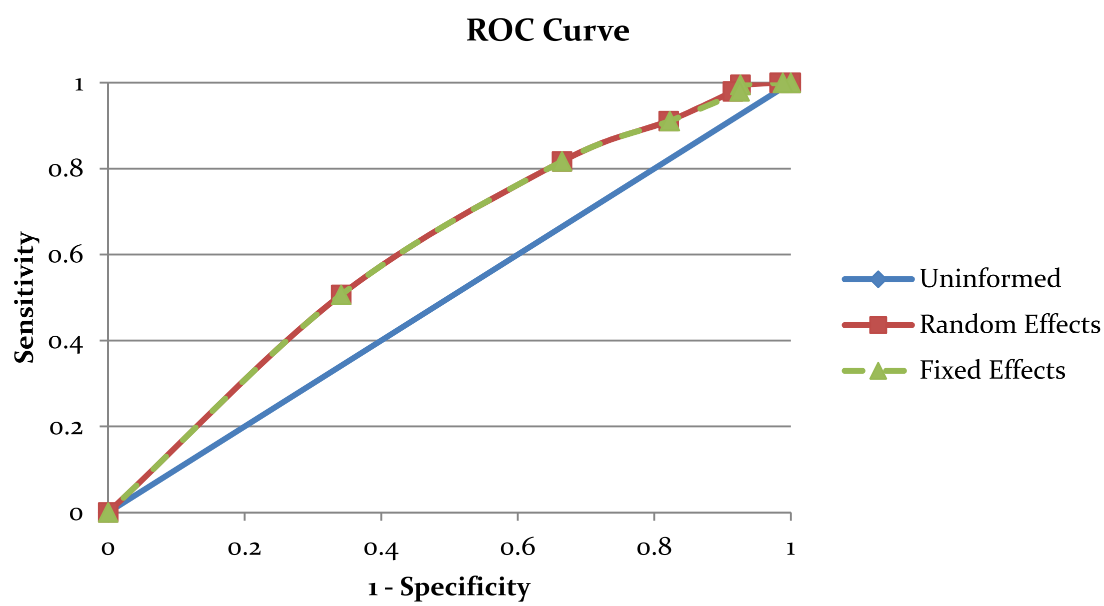
Logit and probit models combine the individual effects of the variables; however a panel approach may be able to map the complexity of financial markets much more efficiently. Therefore, logit and probit models using random and fixed effects were estimated. The results for binary panel estimates confirm the findings of Bussiere and Fratzscher (2006) as they do not add value to binary forecasts.

4.5.2 Panel Estimations

Binary modeling may not be market players' weapon of choice as it has limited ability to forecast 1-month ahead currency movements since the binary crash definition truncates information. Instead of predicting whether or not a crash will materialize, one could be more interested in both the direction and the magnitude of the change in a currency pair. For this reason, percentage returns were regressed on the same variables in a panel setting. The results for random and fixed effects models are presented in Appendix P.

The results indicate that the panel approaches capture the FX market fluctuations well. This is possibly due to the inclusion of macroeconomic variables, namely, interest rate premium, current account balance and growth rate. The adjusted ROC curve for the panel models, presented in Figure 4.4, produces an AUC of 0.6090 and 0.6084 for random and fixed effects models, respectively. The adjusted ROC curve is of importance as it encompasses potential gain and loss information. An AUC greater than 0.5 for an adjusted ROC curve indicates that even a simple trading strategy of buying and holding until the end of each month would produce financial profits.

Figure 4.4: Adjusted ROC Curve for Random and Fixed Effects Models



Appendix P shows that both random and fixed effects models employed debt related macroeconomic variables as well as market variables, VIX and TED-Spread. Figure 4.4, on the other hand clearly shows that point estimations form by using random or fixed effects models produce meaningful financial profits as the AUC for both panel models are well above the 0.5 mark. Consequently, panel models not only produce good directional forecasts but also predict these directions with acceptable opportunity costs. Hence, it can be concluded that, the joint use of macroeconomic variables with market variables in a panel environment is behind the success of point estimations.

As a final addition to all models used in this chapter, lagged crash identifiers and lagged returns were added to binary and panel models, respectively. However, unlike Candelon, Dumitrescu and Hurlin (2014), both lagged variables were found to contain no additional information about 1-month ahead currency movements which is in line with weak form market efficiency. Therefore, the discrepancy between the findings of this chapter and Candelon, Dumitrescu and Hurlin (2014) may be due to their crisis definition or the use of emerging markets.

The results presented in this section show that we are still a long way from an omniscient EWS. Informed prediction of currency movements in some currency pairs on a country-by-country basis is just harder than others. Such is the case for Switzerland, Canada, Japan and New Zealand. However, there may still be explanatory power to be gained from data aggregation since results for panel approaches produce profitable FX strategies. Furthermore, the lack of past lagged crash indicators and lagged returns in any one of the models indicates that currency markets process shocks reasonably fast.

4.6 Discussion of Results

The results for the multiple linear regression models on the other hand indicate weak explanatory power for most macroeconomic variables. Though these variables may be expected to hold predictive power in longer forecast horizons, the 1-month forecasting horizon used in this chapter may have reduced their power.

The indicator approach, using a simple crash definition of 2% loss, is able to produce informed predictions for 1-month ahead currency crashes. However, the number of signals required to indicate a crash has a sizeable effect on the performance of the signaling approach. Hence, the number of signals required to indicate a crash must be increased to produce a satisfactory binary model. Nonetheless, the indicator approach is quite successful when using 3 (4) or more signals to indicate a crash with classification accuracy well above 0.60 (0.70).

Similar to the results obtained in the regression analysis, logit and probit models find most macroeconomic variables to be insignificant. In contrast to macroeconomic variables, market variables VIX and TED-Spread are consistently significant in both binary and point estimate models. The difference between the ROC curves for country-by-country binary estimations and pooled binary estimations points to gains from pooling. Furthermore, I find evidence against dynamic specifications for binary models as lagged crash indicators turn out to be consistently insignificant. However, the binary performance of logit and probit models, especially when making country-by-country predictions, are less than satisfactory as they are incapable of producing any crisis predictions for Switzerland, Canada, Japan and New Zealand. Hence, the results presented in this chapter support signaling models rather than logit and probit for binary crash prediction.

On the other hand, panel approaches seem to have captured the currency market dynamics better. This may be due to the inclusion of three macroeconomic variables, namely, interest rate premium, current account balance and growth rate. Nonetheless, the use of these variables does not render market variables VIX and TED-Spread insignificant. The adjusted AUC for both fixed and random effects models are well above 0.5, which indicates panel models using market variables are capable of generating financial profits. Moreover, similar to binary models, lagged returns are found to be insignificant for panel models, ruling out dynamic panel approaches.

The macroeconomic variables found to be significant in the panel approach highlight an important feature of currency markets. Interest rate premium, current account balance and growth rate are all directly related to the debt servicing ability of a country. Hence, the addition of these variables to market variables such as TED-Spread or VIX in a panel setting is in line with theory (Krugman (2002; 2010)). Given the performance of panel models, it may be argued that banks' asset-side problems affected global markets relatively fast. Hence, despite the lack of a direct asset-bubble parameter, the panel models were able to provide successful investment decisions by taking into account market variables. However, the lack of a direct asset bubble indicator, the missing link, still presents a shortcoming for EWS. Thus, forming an asset bubble indicator must be the next step in creating next generation EWS.

Several binary and panel estimation models were reviewed in this chapter using both macroeconomic variables and market variables as predictors. Market variables, VIX and TED-Spread, are not only new to the currency crash literature but also help us better understand how currency markets operate. Much like market heat predicting high frequency crashes via liquidity related order book variables, the models tested in this chapter successfully forecast several currency crises using liquidity based market variables. Hence, Chapter 4 reaffirms the need to focus on market players' perception of liquidity risk when making both short and long term crisis predictions.

The work presented in this chapter contributes to the EWS literature by successfully predicting currency crashes and completes the second of goal of this thesis, crash prediction. Specifically, Chapter 4 contributes to the literature by creating both binary and panel prediction models to successful forecast currency crashes for the seldom studied G10 countries. Furthermore, this chapter contributes to the literature by introducing market variables, which are found to carry significant crash related information, into the EWS literature. Perhaps the most important contribution of this chapter however is in its ability to produce meaningful profits using the point estimates from panel models. The successful 1-month currency predictions formed here may be useful to investors as the prediction period is much shorter compared to most EWS studies that use 12 to 24 month horizons.

Chapter 5: Conclusion

5.1 Summary of Work

The two primary goals of this thesis were to recover the normal distribution assumption for high frequency returns and to predict financial crashes that occur in both the short and the long run. While Chapter 2 focused on fulfilling the first research goal by obtaining normally distributed high frequency returns via subordination, it also identified the fundamental elements that influence the price formation process.

In Chapter 2, a novel way to look at high frequency returns, natural time, was introduced. Instead of calendar time, natural time approach sampled the high frequency data using transaction time and subordinated the raw returns with order book variables. Essentially, natural time corrects for the deviations from normality, often observed when working with high frequency data. In other words, natural time adjusts for the true flow of time, and hence information, in financial markets by sampling under transaction time and by employing order book variables in stochastic subordination. Consequently, the natural time approach corrects the data for heteroscedasticity.

The top 10 most liquid stocks traded at the LSE were used to evaluate natural time and on several occasions the natural time approach was able to subordinate returns to arrive at normal distributions. Natural time was also superior to the GARCH model indicating that the subordination functions were better predictors of volatility compared to the benchmark GARCH model.

In Chapter 3, I put to test the applicability of the variables used in the natural time approach to binary flash crash prediction. The results showed that the key variables used to account for non-normality of high frequency returns also contain information about impending flash crashes, even though the variables are formed under calendar time sampling this time. Hence, the order book variables that were used for subordination in Chapter 2 were critical in creating the high frequency crash prediction metric, market heat, in Chapter 3.

Combining linear discriminant analysis with order book variables, MH successfully predicted short term financial crashes, part of the second goal of this research. Specifically, MH combined three critical components of high frequency markets: speed, liquidity and momentum which are captured by order book update duration, liquidity and volume imbalance variables, respectively. MH was tested against a linear and a VPIN-based model. Although all three models were capable of predicting the Flash Crash, MH consistently outperformed all alternatives across different markets and crash thresholds using both in-sample and out-of-sample data. The broad applicability of the order book variables identified in Chapter 2 to events as dramatic as the Flash Crash suggests that in high frequency settings order book information is relevant for both high frequency traders and policy makers. Given its performance, MH could indeed be incorporated into a circuit breaker to avoid future episodes like the Flash Crash.

Flash crashes only make up a portion of the financial turbulences we observe in the finance world. To address the problem of crash prediction in longer time horizons, I shifted the focus to the macroeconomic level in Chapter 4, creating EWS for currency crashes, thus completing the second objective of this thesis.

In Chapter 4, I extend the time scale for crash prediction to predict currency crashes in G10 countries using both binary and panel models. The binary models tested include the indicator approach, logit and probit while fixed and random effects models were used to form point estimates.

Naturally, as one moves from 5-minute ahead forecasts, such as those produced by MH, to 1-month ahead predictions, the state of the economy and hence macroeconomic variables become relevant in determining the value of the currency. Hence, macroeconomic variables were tested along with market variables.

One important finding regarding all models used in Chapter 4 is that market variables VIX and TED-Spread were always found to be significant. While logit and probit model result were not satisfactory, the signaling model was quite successful. Both panel models shared three macroeconomic variables, namely, interest rate premium, current account balance and growth rate. This shows that for point estimates macroeconomic variables directly related to the debt servicing ability of a country are of importance.

Chapter 4 augmented the signaling model of KLR and introduced a working binary currency crash predictor. The panel estimations were also very successful. So much so that the adjusted AUC values for both fixed and random effects models were above 0.60 indicating profitable buy-and-hold strategies.

The performance of the indicator approach and the panel models and the lack of an asset-side variable, suggest that information regarding asset-side problems were fed quickly to liquidity and investor sentiment based market variables. Thus, VIX and TED-Spread were good approximators for a direct asset-side indicator. Nonetheless, the need for an asset-side indicator still remains a shortcoming for all EWS which aim to capture the imbalances in balance sheets.

5.2 Contributions

This thesis contributes to several strands of finance literature. Natural time, a novel subordination procedure introduced in Chapter 2, contributes to the subordination literature by successfully achieving normal return distributions on several occasions. Natural time adds to the growing evidence that asset returns are normally distributed even in the high frequency, as long as one accounts for the latent process returns are subordinated to. Natural time approach samples the transaction data as they occur in tick time using a composite index of volume, liquidity and duration as the subordinator. The success of natural time indicates that the culprit for the non-normal distribution of financial returns is the imposition of a time grid by sampling data in calendar time coupled with not accounting for the information contained within the order book.

The contribution of natural time is threefold. First contribution of natural time is that it is the first study to successful subordinate returns with order book variables under transaction time. Second contribution of natural time is in its subordination variables. By extending the range of order book variables used in subordination and employing a nonlinear asymmetric response function, natural time accounts for the underlying information process much more efficiently. Finally, given its consistent superiority to GARCH, natural time contributes to the literature by creating a superior volatility gauge that accounts for the heteroscedasticity in the data better than the ubiquitous GARCH model.

All in all, natural time reconciles the empirical observation of financial returns with finance theory by recovering one of the central assumptions of finance, normal distribution.

Market heat, on the other hand, contributes to high frequency crash prediction literature by proving that flash crashes are predictable, both for single stocks and indices. The excellent performance of MH rests on two pillars: the use of liquidity related order book variables and perfect classification of trades.

In a binary crash prediction setting using E-Mini S&P500 futures data, market heat is found to be superior in classification accuracy, precision and recall compared to all alternative methods tested. The findings are robust across different timeframes and crash thresholds and extend to several single stocks traded on the London Stock Exchange. Hence, market heat contributes to the growing flash crash literature by offering a robust flash crash prediction tool where liquidity is the key driver of high frequency returns. MH could also be utilized as a circuit breaker by stock exchanges to avoid future flash crash episodes.

Additionally, unlike most existing literature, natural time and market heat classify high frequency transactions perfectly into buyer or seller initiated trades via codes provided by the stock exchange or order submission times. In this respect, this thesis establishes a best practice for any high frequency study by classifying trades perfectly, an approach that should always be preferred to outdated bulk classification methods invented before the widespread availability of order book data.

In Chapter 4, the crash prediction framework is extended to include low frequency macroeconomic currency crises. Chapter 4 contributes to the early warning system literature by developing successful EWS for developed markets instead of the often studied emerging markets and offers evidence on the predictability of currency crashes.

In order to capture both asset and liability related crashes, market variables, VIX and TED-Spread were used in addition to macroeconomic indicators, a unique contribution of this work to the EWS literature. Using a crash threshold of 2% loss and a 1-month forecast horizon, most macroeconomic variables are found to be insignificant in binary models while market variables, VIX and TED-Spread, add considerable forecasting power to both binary and panel models. Macroeconomic variables related to debt servicing are also found to have explanatory power for panel estimations.

Chapter 4 contributes to the binary currency crash literature by introducing a working binary indicator of currency crashes for G10 countries, the signaling model with 3 or more signals to indicate a crash. The most important contribution of Chapter 4 however lies in its ability to generate profits via the point estimates generated by panel models that combine macroeconomic and market variables.

Despite the lack of a direct asset side variable, the success of both the indicator approach and the panel models indicate that the market variables used in this study were able to effectively account for asset side problems. In fact, the almost complete absence of macroeconomic variables in logit and probit estimations and the central role of liquidity based market variables in determining currency fluctuations suggest that liquidity concerns govern asset returns both in the short and the long run. Thus, the cardinal contribution of this thesis is the solid evidence it provides for liquidity as the chief determinant of asset price depreciations in both the short and the long term.

5.3 Future Work

Natural time proves that the use of order book variables is effective in accounting for the latent price process. Future subordination based studies can benefit from this finding by including an array of additional order book variables to capture information about high frequency returns.

Furthermore, given natural time's superiority to GARCH, the information advantage gained from using order book data could also be used to form a volatility forecasting tool. Although market heat used some of the order book variables to forecast crashes, the ultimate evaluation of the usefulness of order book variables would be possible by direct assessment of these variables as variance predictors. However, as the main goal of natural time was to recover normal distribution of high frequency returns, volatility related use of order book variables stands as a future venue of research.

Binary classification, especially in heavily imbalanced sets such as the one used to test market heat, leaves researchers with a distinct dilemma: high recall or high precision. Market heat primarily focused on generating a signal each time a crash event was to occur. Hence, market heat rarely missed a crash, although it sometimes produced a high number of false positives.

For a market player who wishes to use MH to generate profits, the financial gain of shorting a stock or an index each time MH produces a signal should be quantified before MH is used as a trading rule. Similarly, for a stock exchange whose primary duty is to ensure an orderly and efficient market, the low precision of MH may prove to be problematic as the efficiency cost of halting markets, when no crisis is bound to happen, may be unacceptable. In such instances, stock markets may still combine the order book variables used in MH with alternative binary prediction techniques where more weight is given to obtaining a higher precision value. These two alternative uses of MH surely constitute valid options for the future research.

The adjusted ROC curves for the panel approaches tested in Chapter 4, point to possible financial gains even with a simple buy and hold strategy. The alpha generation capacity of these models needs to be tested taking into account execution costs. The back-testing of such trading strategies would certainly add to the validity of market variables as proxies for asset side problems in EWS. Such trade based assessment of market variables are left for future research.

Bibliography

Abiad, Abdul. 2003. Early-warning systems: A Survey and A Regime-Switching Approach. IMF Working Paper WP/03/32, International Monetary Fund.

Admati, Anat R., & Pfleiderer, Paul. 1988. A Theory of Intraday Patterns: Volume and Price Variability. Review of Financial Studies, **1 (1)**, 3–40.

Aït-Sahalia, Yacine, & Yu, Jialin. 2009. High Frequency Market Microstructure Noise Estimates and Liquidity Measures. The Annals of Applied Statistics, **3 (1)**, 422–457.

Aït-Sahalia, Yacine, Mykland, Per A., & Zhang, Lan. 2005. How Often To Sample A Continuous-Time Process in the Presence of Market Microstructure Noise. Review of Financial Studies, **18 (2)**, 351–416.

Aït-Sahalia, Yacine, Fan, Jianqing, & Xiu, Dacheng. 2010. High-Frequency Covariance Estimates With Noisy and Asynchronous Financial Data. Journal of the American Statistical Association, **105 (492)**, 1504–1517.

Aldridge, Irene. 2010. High-Frequency Trading : A Practical Guide to Algorithmic Strategies and Trading Systems. New Jersey: Wiley.

Allen, David, Lazarov, Zdravetz, McAleer, Michael, & Peiris, Shelton. 2009. Comparison of Alternative ACD Models via Density and Interval Forecasts: Evidence from the Australian Stock Market. Mathematics and Computers in Simulation, **79 (8),** 2535-2555.

Almgren, Robert, & Chriss, Neil. 2000. Optimal Execution of Portfolio Transactions. Journal of Risk, **3 (2)**, 5–39.

Amihud, Yakov, & Mendelson, Haim. 2000. The Liquidity Route to a Lower Cost of Capital. Journal of Applied Corporate Finance, **12 (4)**, 8–25.

Anand, Amber, Chakravarty, Sugato, & Martell, Terrence. 2005. Empirical Evidence on the Evolution of Liquidity: Choice of Market Versus Limit Orders by Informed and Uninformed Traders. Journal of Financial Markets, **8 (3)**, 289–309.

Andersen, Torben G., & Bollerslev, Tim. 1997. Intraday Periodicity and Volatility Persistence in Financial Markets. Journal of Empirical Finance, **4 (2-3)**, 115–158.

Andersen, Torben G., Bollerslev, Tim, Diebold, Francis X., & Labys, Paul. 2001. The Distribution of Realized Exchange Rate Volatility. Journal of the American Statistical Association, **96 (453)**, 42–55.

Andersen, Torben G., Bollerslev, Tim, Diebold, Francis X., & Vega, Clara. 2003. Micro Effects of Macro Announcements: Real-Time Price Discovery in Foreign Exchange. The American Economic Review, **93 (1)**, 38–62.

Andersen, Torben G., Bollerslev, Tim, Diebold, Francis X., & Vega, Clara. 2007. Real Time Price Discovery in Global Stock, Bond and Foreign Exchange Markets. Journal of International Economics, **73 (2)**, 251–277.

Andersen, Torben G. and Bondarenko, Oleg. 2014a. VPIN and the Flash Crash. Journal of Financial Markets, **17 (1)**, 1-46.

Andersen, Torben G. and Bondarenko, Oleg. 2014b. Reflecting on the VPIN Dispute. Journal of Financial Markets, **17 (1)**, 53-64.

Andersen, Torben G. and Bondarenko, Oleg. 2015. Assessing Measures of Order Flow Toxicity and Early Warning Signals for Market Turbulence. Review of Finance, **19 (1)**, 1-54.

Ané, Thierry, & Geman, Hélyette. 2000. Order Flow, Transaction Clock, and Normality of Asset Returns. Journal of Finance, **55 (5)**, 2259–2284.

Apolinario, Rosa María Cáceres, Santana, Octavio Maroto, Sales, Lourdes Jordán, & Caro, Alejandro Rodríguez. 2006. Day of the Week Effect on European Stock Markets. International Research Journal of Finance and Economics, **1 (2)**, 53–70.

Bachelier, Louis. 1900. Théorie de la Spéculation. Annales Scientifiques de L'école Normale Supérieure, **3 (17)**, 21–86.

Bagehot, Walter. 1971. The Only Game In Town. Financial Analysts Journal, **27 (2)**, 12–14.

Bandi, Federico M., & Russell, Jeffrey R. 2008. Microstructure Noise, Realized Variance, and Optimal Sampling. Review of Economic Studies, **75 (2)**, 339–369.

Barclay, Michael J., & Warner, Jerold B. 1993. Stealth Trading and Volatility: Which Trades Move Prices? Journal of Financial Economics, **34 (3)**, 281–305.

Barndorff-Nielsen, Ole E., Hansen, Peter Reinhard, Lunde, Asger, & Shephard, Neil. 2011a. Subsampling Realised Kernels. Journal of Econometrics, **160 (1)**, 204-219.

Barndorff-Nielsen, Ole E., Hansen, Peter R., Lunde, Asger, & Shephard, Neil. 2011b. Multivariate Realized Kernels: Consistent Positive Semidefinite Estimators of the Covariation of Equity Prices with Noise and Nonsynchronous Trading. Journal of Econometrics, **162** **(2)**, 149–169.

Beckmann, Daniela, Menkhoff, Lukas, & Sawischlewski, Katja. 2006. Robust Lessons About Practical Early Warning Systems. Journal of Policy Modeling, **28** **(2)**, 163–193.

Berg, Andrew, & Pattillo, Catherine. 1999. Predicting currency crises: The Indicators Approach and An Alternative. Journal of International Money and Finance, **18** **(4)**, 561–586.

van den Berg, Jeroen, Candelon, Bertrand, and Urbain, Jean-Pierre. (2008). A Cautious Note on the Use of Panel Models to Predict Financial Crises, Economics Letters, **101** (**1**), 80-83.

Bertsimas, Dimitris, & Lo, Andrew W. 1998. Optimal Control of Execution Costs. Journal of Financial Markets, **1** **(1)**, 1–50.

Biais, Bruno, Glosten, Larry, & Spatt, Chester. 2005. Market Microstructure: A Survey of Microfoundations, Empirical Results, and Policy Implications. Journal of Financial Markets, **8** **(2)**, 217–264.

Black, Fischer, & Scholes, Myron. 1973. The Pricing of Options and Corporate Liabilities. Journal of Political Economy, **81** **(3)**, 637–654.

Blau, Benjamin M. 2009. Intraday Stealth Trading: Which Trades Move Prices During Periods of High Volume. The Journal of Financial Research, **32 (1)**, 1-21.

Bollerslev, Tim. 1986. Generalized Autoregressive Conditional Heteroscedasticity. Journal of Econometrics, **31 (3)**, 307–327.

Brunnermeier, Markus K., Nagel, Stefan, & Pedersen, Lasse H. 2008. Carry Trades and Currency Crashes. NBER Macroeconomics Annual 2008, **23 (1)**, 313-347.

Brüggemann, Axel, & Linne, Thomas. 2002. Are the Central and Eastern European Transition Countries Still Vulnerable to a Financial Crisis? Results from the Signals Approach. BOFIT Discussion Papers No. 5, Bank of Finland, Institute for Economies in Transition.

Bussiere, Matthieu, & Fratzscher, Marcel. 2006. Towards a New Early Warning System of Financial Crises. Journal of International Money and Finance, **25 (6)**, 953-973.

Cai, Jinghan, Ouyang, Hongbing & Wong, Michael C.S. 2011. The Bear Market in China: Which Trades Push The Stock Prices Down. Annals of Financial Economics, **6 (1)**, 1-22.

Candelon, Bertrand, Dumitrescu, Elena-Ivona, & Hurlin, Christophe. 2014. Currency Crises Early Warning Systems: Why They Should Be Dynamic. International Journal of Forecasting, 30 (4), 1016-1029.

CFTC-SEC. 2010. Preliminary Findings Regarding the Market Events of May 6, 2010. Joint Commodity Futures Trading Commission (CFTC) and the Securities and Exchange Commission (SEC) Advisory Committee on Emerging Regulatory Issues, May 18, 2010.

Chakrabarty, Bidisha, Pascual, Roberto, and Shkilko, Andriy. 2013. Trade Classification Algorithms: A Horse Race Between the Bulk and the Tick-based Rules. The University of the Balearic Islands.
Available at: <http://dee.uib.es/digitalAssets/234/234006_Pascual.pdf>.

Chakravarty, Sugato. 2001. Stealth-Trading: Which Traders' Trades Move Stock Prices? Journal of Financial Economics, **61 (2)**, 289–307.

Chang, Roberto, & Velasco, Andrés. 2000. Financial Fragility and the Exchange Rate Regime. Journal of Economic Theory, **92 (1)**, 1–34.

Ciarlone, Alessio, & Trebeschi, Giorgio. 2005. Designing An Early Warning System for Debt Crises. Emerging Markets Review, **6 (4)**, 376–395.

Clark, Peter K. 1973. A Subordinated Stochastic Process Model with Finite Variance for Speculative Prices. Econometrica, **41 (1)**, 135–155.

Cont, Rama. 2001. Empirical Properties of Asset Returns: Stylized Facts and Statistical Issues. Quantitative Finance, **1 (2)**, 223–236.

Corsetti, Giancarlo, Pesenti, Paolo, & Roubini, Nouriel. 1999. Paper Tigers?: A Model of the Asian Crisis. European Economic Review, **43 (7)**, 1211–1236.

Cumperayot, Phornchanok, & Kouwenberg, Roy. 2013. Early Warning Systems for Currency Crises: A Multivariate Extreme Value Approach. Journal of International Money and Finance, **36 (1)**, 151-171.

Dacorogna, Michel M., Müller, Ulrich A., Nagler, Robert J., Olsen, Richard B., & Pictet, Olivier V. 1993. A Geographical Model for the Daily and Weekly Seasonal Volatility in the Foreign Exchange Market. Journal of International Money and Finance, **12 (4)**, 413–438.

Dacorogna, Michel M., Gaureau, Cindy L., Müller, Ulrich A., Olsen, Richard B., & Pictet, Olivier V. 1996. Changing Time Scale for Short-Term Forecasting in Financial Markets. Journal of Forecasting, **15 (3)**, 203–227.

Dacorogna, Michel M., Gençay, Ramazan, Müller, Ulrich A., Olsen, Richard B., & Pictet, Olivier V. 2001. An Introduction to High-Frequency Finance. San Diego: Academic Press.

Davis, E. Philip, & Karim, Dilruba. 2008. Comparing Early Warning Systems for Banking Crisis. Journal of Financial Stability, **4 (2)**, 89–120.

deB. Harris, Fredrick H., McInish, Thomas H., Shoesmith, Gary L., & Wood, Robert A. 1995. Cointegration, Error Correction, and Price Discovery on Informationally Linked Security Markets. Journal of Financial and Quantitative Analysis, **30 (4)**, 563–579.

Demirgüç-Kunt, Aslı, & Detragiache, Enrica. 1998. The Determinants of Banking Crises in Developing and Developed Countries. IMF Staff Papers, **45 (1)**, 81-109.

Deuskar, Prachi and Johnson, Timothy C. 2014. Market Liquidity and Flow-driven Risk. The Review of Financial Studies, **24 (3)**, 721-753.

Do, Hung X., Brooks, Robert, Treepongkaruna, Sirimon & Wu, Eliza. 2014. How Does Trading Volume Affect Financial Return Distributions. International Review of Financial Analysis, **35 (1)**, 190-206.

Easley, David, Engle, Robert F., O'Hara, Maureen, & Wu, Liuren. 2008. Time-Varying Arrival Rates of Informed and Uninformed Trades. Journal of Financial Econometrics, **6 (2)**, 171-207.

Easley, David, Lopez de Prado, Marcos and O'Hara, Maureen. 2011a. The Exchange of Flow Toxicity. The Journal of Trading, **6 (2)**, 8-13.

Easley, David, Lopez de Prado, Marcos and O'Hara, Maureen. 2011b. The Microstructure of the 'Flash Crash': Flow Toxicity, Liquidity Crashes and the Probability of Informed Trading. The Journal of Portfolio Management, **37 (2)**, 118-128.

Easley, David, Lopez de Prado, Marcos and O'Hara, Maureen. 2012. Flow Toxicity and Liquidity in a High Frequency World. Review of Financial Studies, **25 (5)**, 1457-1493.

Easley, David, Lopez de Prado, Marcos and O'Hara, Maureen. 2015. Discerning Information from Trade Data. Johnson School Research Paper Series No. 8-2012. Available at: <http:/ssrn.com/abstract=1989555>.

Easley, David, Lopez de Prado, Marcos and O'Hara, Maureen. 2014. VPIN and the Flash Crash: A Rejoinder. Journal of Financial Markets, **17 (1)**, 47-52.

Easley, David, & O'Hara, Maureen. 1987. Price, Trade Size, and Information in Securities Markets. Journal of Financial Economics, **19 (1)**, 69–90.

Easley, David, & O'Hara, Maureen. 1992. Time and the Process of Security Price Adjustment. Journal of Finance, **47 (2)**, 577–605.

Edison, Hali J. 2000. Do Indicators of Financial Crises Work? An Evaluation of an Early Warning System. International Finance Discussion Papers No. 675, Board of Governors of the Federal Reserve System.

Eichengreen, Barry, Rose, Andrew K., & Wyplosz, Charles. 1996. Contagious Currency Crises. Working Paper No. 5681, National Bureau of Economic Research. Available at: < http://ssrn.com/abstract=4281 >

Engle, Robert F. 1982. Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation. Econometrica, **50 (4)**, 987–1007.

Engle, Robert F., & Russell, Jeffrey R. 1998. Autoregressive Conditional Duration: A New Model for Irregularly Spaced Transaction Data. Econometrica, **66 (5)**, 1127–1162.

Entorf, Horst, Gross, Anne, & Steiner, Christian. 2009. Return and Volatility Reactions to Monthly Announcements of Business Cycle Forecasts: An Event Study Based on High Frequency Data. ZEW Discussion Papers No. 09-010, ZEW Zentrum für Europäische Wirtschaftsforschung / Center for European Economic Research.

Epps, Thomas W. 1979. Comovements in Stock Prices in the Very Short Run. Journal of American Statistical Association, **74 (366)**, 291–298.

Esquivel, Gerardo, & Felipe, Larraín B. 1998. Explaining Currency Crises. John F. Kennedy Faculty Research WP Series R98-07, El Colegio de Mexico, Centro de Estudios Economico, Pontificia Universidad Catolica de Chile.

Fama, Eugene F. 1965. The Behavior of Stock Market Prices. Journal of Business, **38 (1)**, 34–105.

Fisher, Ronald A. 1936. The Use of Multiple Measurements in Taxonomic Problems. Annals of Eugenics, **7 (2)**, 179-188.

French, Kenneth R. 1980. Stock Returns and the Weekend Effect. Journal of Financial Economics, **8 (1)**, 55–69.

Geman, Hélyette. 2002. Pure jump Lévy Processes for Asset Price Modelling. Journal of Banking & Finance, **26 (7)**, 1297–1316.

Geman, Hélyette. 2005. From Measure Changes to Time Changes in Asset Pricing. Journal of Banking & Finance, **29 (11)**, 2701–2722.

Gibbons, Michael R., & Hess, Patrick. 1981. Day of the Week Effects and Asset Returns. Journal of Business, **54 (4)**, 579–596.

Gillemot, László, Farmer, J. Doyne, & Lillo, Fabrizio. 2006. There's More to Volatility Than Volume. Quantitative Finance, **6 (5)**, 371–384.

Glattfelder, James B., Dupuis, Alexandre, & Olsen, Richard B. 2011. Patterns in High Frequency FX Data: Discovery of 12 Empirical Scaling Laws. Quantitative Finance, **11 (4)**, 599–614.

Glosten, Lawrence R., & Milgrom, Paul R. 1985. Bid, Ask and Transaction Prices in a Specialist Market with Heterogeneously Informed Traders. Journal of Financial Economics, **14 (1)**, 71–100.

Golub, Anton, Keane, John and Poon, Ser-Huang. 2012. High-Frequency Trading and Mini Flash Crashes. Working paper, University of Manchester. Available at: <http://ssrn.com/abstract=2182097>.

Hansen, Peter R., & Lunde, Asger. 2006. Realized Variance and Market Microstructure Noise. Journal of Business & Economic Statistics, **24 (2)**, 127–161.

Harada, Kimie, & Watanabe, Toshiaki. 2009. News Effects on High Frequency Yen/Dollar Exchange Rates and Its Volatility Behavior. Working paper, Chuo University.

Hautsch, Nikolaus, Hess, Dieter, & Veredas, David. 2011. The Impact of Macroeconomic News on Quote Adjustments, Noise, and Informational Volatility. Journal of Banking and Finance, **35 (10)**, 2733–2746.

Hawkins, John, & Klau, Marc. 2000. Measuring Potential Vulnerabilities in Emerging Market Economies. BIS Working Paper No. 91, Bank of International Settlements. Available at: < http://ssrn.com/abstract=2182097>.

Huang, Han-Ching. 2011. An Analysis of Intraday Return – Order Imbalance Relation to Stealth Trading. Investment Management and Financial Innovation, **8 (1)**, 234-242.

Huang, Huiyi & Lee, Tae-Hwy. 2013. Forecasting Realized Volatility Using Subsample Averaging. Open Journal of Statistics, **3 (5)**, 379-383.

Huth, Nicolas & Abergel, Frederic. 2012. The Times Change: Multivariate Subordination. Empirical Facts. Quantitative Finance, **12(1)**, 1-10.

Instefjord, Norvald. 2005. Investment management. University of London Press.

Jordá, Óscar, & Taylor, Alan M. 2012. The Carry Trade and Fundamentals: Nothing to Fear but FEER Itself. Journal of International Economics, **88 (1)**, 74-90.

Jurek, Jakub W. 2014. Crash-Neutral Currency Carry Trades. Journal of Financial Economics, **113 (3)**, 325-347.

Kaminsky, Garciela, Lizondo, Saul, & Reinhart, Carmen M. 1998. Leading Indicators of Currency Crises. IMF Staff Papers, **45 (1)**, 1-48.

Kapetanios, George. 2003. Determining the Poolability of Individual Series in Panel Datasets. Working Paper No. 499, Queen Mary University of London.

Karpoff, Jonathan M. 1987. The Relation Between Price Changes and Trading Volume: A Survey. Journal of Financial and Quantitative Analysis, **22 (1)**, 109–126.

Kauppi, Heikki, & Saikkonen, Pentti. 2008. Predicting U.S. Recessions with Dynamic Binary Response Models. Review of Economic Studies, **90 (4)**, 777–791.

Kindman, Andrew. 2010. Currency Crises Early Warning Systems: Robust Adjustments to the Signal-Based Approach. Ph.D Thesis. Duke University.

Kirilenko, Andrei, Kyle, Albert S., Samadi, Mehrdad and Tüzün, Tuğkan. 2014. The Flash Crash: The Impact of High Frequency Trading on an Electronic Market. NBER Working Paper. Available at: <http://ssrn.com/abstract=1686004>

Krugman, Paul. 1979. A Model of Balance-of-Payments Crises. Journal of Money, Credit, and Banking, **11 (3)**, 311–325.

Krugman, Paul. 1999. Balance Sheets, the Transfer Problem, and Financial Crises. International Tax and Public Finance, **6 (4)**, 459–472.

Krugman, Paul. 2002. Crises: The Next Generation. Chapter from Economic Policy in the International Economy: Essays in Honor of Assaf Razin, Cambridge University.

Krugman, Paul. 2010. Crises. Nobel Lecture at the Annual Meeting of the American Economic Association.

Kumar, Mohan, Moorthy, Uma, & Perraudin, William. 2003. Predicting Emerging Market Currency Crashes. Journal of Empirical Finance, **10 (4)**, 427–454.

Kyle, Albert S. 1985. Continuous Auctions and Insider Trading. Econometrica, **53 (6)**, 1315–1335.

Large, Jeremy. 2007. Measuring the resiliency of an electronic limit order book. Journal of Financial Markets, **10 (1)**, 1–25.

Lee, Charles M.C & Ready, Marc J. 1991. Inferring Trade Direction from Intraday Data. Journal of Finance, **46 (2)**, 733-746.

Lo, Stephanie, & Rogoff, Vincent R. 2014. Secular Stagnation, Debt Overhang and Other Rationales for Sluggish Growth, Six Years On. Proceedings from the 13th Annual BIS Conference, Lucerne, Switzerland, June 27, 2014.

Lundin, Mark C., Dacorogna, Michel M., & Müller, Ulrich A. 1998. Correlation of High Frequency Financial Time Series. In: Lequeux, Pierre (ed), The Financial Markets Tick by Tick. London: John Wiley & Sons.

Malik, Azeem, & Ng, Wing Lon. 2009. Detecting Diurnal Volume Profile: An Empirical Analysis of Market Depth. Banking and Finance Letters, **1 (3)**, 119–128.

Mandelbrot, Benoît B. 1963. The Variation of Certain Speculative Prices. Journal of Business, **36 (4)**, 394–419.

Mandelbrot, Benoît B., & Taylor, Howard M. 1967. On the Distribution of Stock Price Differences. Operations Research, **15 (6)**, 1057–1062.

Marghescu, Dorina, Sarlin, Peter, & Liu, Shuhua. 2010. Early Warning Analysis for Currency Crises in Emerging Markets: A Revisit with Fuzzy Clustering. Intelligent Systems in Accounting, Finance and Management, **17 (3-4)**, 143–165.

Martens, Martin, Chang, Yuan-Chen, & Taylor, Stephen J. 2002. A Comparison of Seasonal Adjustment Methods When Forecasting Intraday Volatility. The Journal of Financial Research, **25 (2)**, 283-299.

Merton, Robert C. 1976. Option Pricing When Underlying Stock Returns Are Discontinuous. Journal of Financial Economics, **3 (1-2)**, 125–144.

Mian, Atif, & Sufi, Amir. 2014. What Explains the 2007-2009 Drop in Employment? Econometrica, **82 (6)**, 2197-2223.

Monroe, Itrel. 1978. Processes That Can Be Embedded in Brownian Motion. The Annals of Probability, **6 (1)**, 42–56.

Moreno, Ramon, & Trehan, Bharat. 2000. Common Shocks and Currency Crises. Working Paper No. 05, Federal Reserve Bank of San Francisco.

Murphy, Anthony, & Izzeldin, Marwan. 2006. Order Flow, Transaction Clock, and Normality of Asset Returns: Ané and Geman (2000) Revisited. Working Paper, Oxford University / Lancaster University.

Müller, Ulrich A., Dacorogna, Michel M., Olsen, Richard B., Pictet, Olivier V., Schwarz, Matthias, & Morgenegg, Claude. 1990. Statistical Study of Foreign Exchange Rates: Empirical Evidence of A Price Change Scaling Law, and Intraday Analysis. Journal of Banking & Finance, **14 (6)**, 1189–1208.

Münnix, Michael C., Schäfer, Rudi, & Guhr, Thomas. 2010. Impact of the Tick-Size on Financial Returns and Correlations. Physica A, **389 (21)**, 4828–4843.

Nag, Ashok K., & Mitra, Amit. 1999. Neural Networks and Early Warning Indicators of Currency Crisis. Reserve Bank of India Occasional Papers, **20 (2)**, 183–222.

Ng,Wing Lon. 2008. Analysing Liquidity and Absorption Limits of Electronic Markets with Volume Durations. Quantitative Finance, **8 (4)**, 353–361.

Obstfeld, Maurice. 1986. Rational and Self-Fulfilling Balance-of-Payments Crises. American Economic Review, **76 (1)**, 72–81.

Obstfeld, Maurice. 1996. Models of Currency Crises with Self-Fulfilling Features. European Economic Review, **40 (3-5)**, 1037–1047.

Oomen, Roel C. A. 2006. Properties of Realized Variance under Alternative Sampling Schemes. Journal of Business & Economic Statistics, **24 (2)**, 219–237.

Osler, Carol L. 2005. Stop-Loss Order and Price Cascades in Currency Markets. Journal of International Money and Finance, **24 (2)**, 219–241.

Pagel, Izabel Mari, de Jongh, Pieter Juriaan & Venter, Johannes Hendrik. 2007. An Introduction to Realized Volatility. Investment Analysts Journal, **36 (65)**, 47-57.

Peria, Maria Soledad Martinez. 1999. A Regime Switching Approach to Studying Speculative Attacks: A Focus on European Monetary System Crises. Policy Research Working Paper No. 2132, World Bank, Development Research Group.

Perry, Philip R. 1983. More Evidence on the Nature of the Distribution of Security Returns. The Journal of Financial and Quantitative Analysis, **18 (2)**, 211–221.

Reinhart, Carmen, Kaminsky, Graciela, & Goldstein, Morris. 2000. Notes on Contagion. MPRA Paper No. 24569, University of Maryland. Available at: <http://mpra.ub.uni-muenchen.de/24569/1/MPRA_paper_24569.pdf>.

Reinhart, Carmen M., Reinhart, Vincent R., & Rogoff, Kenneth S. 2012. Public Debt Overhangs: Advanced-Economy Episodes since 1800. Journal of Economic Perspectives, **26 (3)**, 69–86.

Reinhart, Carmen M., & Rogoff, Kenneth S. 2011. From Financial Crash to Debt Crisis. American Economic Review. **101 (5)**, 1676–1706.

Roll, Richard. 1984. A Simple Implicit Measure of The Effective Bid-Ask Spread in An Efficient Model. Journal of Finance, **39 (4)**, 1127–1139.

Salant, Stephen W., & Henderson, Dale W. 1978. Market Anticipations of Government Policies and the Price of Gold. Journal of Political Economy, **86 (4)**, 627–648.

Savaşer, Tanseli. 2011. Exchange-Rate Response to Macro News: Through the Lens of Market Microstructure. Journal of International Financial Markets, Institutions & Money, **21 (1)**, 107–126.

Savor, Pavel, &Wilson, Mungo. 2015. Earning Announcements and Systematic Risk. AFA 2012 Chicago Meetings Paper, Temple University / University of Oxford.

Silva, C., and Yakovenko, V. M., 2007. Stochastic Volatility of Financial Markets as the Fluctuating Rate of Trading: An Empirical Study. Physica A, **382 (1)**, 278–285.

Stoll, Hans R. 1978. The Supply of Dealer Services in Securities Markets. Journal of Finance, **33 (4)**, 1133–1151.

Taleb, Nassim N. 2008. The Black Swan: The Impact of the Highly Improbable. London: Allen Lane.

Tauchen, George E., & Pitts, Mark. 1983. The Price Variability-Volume Relationship on Speculative Markets. Econometrica, **51 (2)**, 485–505.

Tóth, Bence, & Kertész, János. 2009. The Epps Effect Revisited. Quantitative Finance, **9 (7)**, 793–802.

Tudela, Merxe. 2004. Explaining Currency Crises: A Duration Model Approach. Journal of International Money and Finance, **23 (5)**, 799–816.

Velasco-Fuentes, Rafael, & Ng, Wing Lon. 2010. Nonlinearities in Stochastic Clocks: Trades and Volume as Subordinators of Electronic Markets. Quantitative Finance, **11 (6)**, 863–881.

Veronesi, Pietro. 1999. Stock Market Overreaction to Bad News in Good Times: A Rational Expectations Equilibrium Model. Review of Financial Studies, **12 (5)**, 975–1007.

Wasserfallen, Walter, & Zimmermann, Heinz. 1985. The Behavior of Intra-Daily Exchange Rates. Journal of Banking and Finance, **9 (1)**, 55–72.

Wiener, Norbert. 1923. Differential Space. Journal of Mathematical Physics, **2 (1)**, 131–174.

Wu, Kesheng, Bethel, Wes E., Gu, Ming, Leinweber, David & Oliver Rübel. 2013. A Big Data Approach to Analyzing Market Volatility. Algorithmic Finance, **2 (3-4)**, 241-267.

Zhang, Lan. 2006. Efficient Estimation of Stochastic Volatility Using Noisy Observations: A Multi-Scale Approach. Bernoulli, **12 (6)**, 1019–1043.

Zhang, Lan. 2011. Estimating Covariation: Epps Effect, Microstructure Noise. Journal of Econometrics, **160 (1)**, 33–47.

Zhang, Lan, Mykland, Per A., & Aït-Sahalia, Yacine. 2005. A Tale of Two Time Scales: Determining Integrated Volatility with Noisy High-Frequency Data. Journal of American Statistical Association, **100 (472)**, 1394–1411.

Appendices

Appendix A

Order Book Reconstruction

The order book is reconstructed from the three main files, namely Order Detail, Order History and Trade Report, which contain all necessary information to reconstruct the state of the order book at any moment. Order Detail files provide information on new orders submitted to the exchange, while Order History files provide information on alteration of previously submitted orders. The types of change that are allowed by the Exchange are deletions, expiries, full or partial matches and transaction limits. Order Detail and Order History files are sufficient to construct the standing order book, as information on addition and removal of each trade that appears on the order book can be found within these files. The Trade Report file, on the other hand, is used to determine trade times which are reported to the nearest second. The sequence of trades, matching orders and the trade initiators can all be determined using the information in this file.

The sequence of orders are determined by the timestamps and the message sequence number (MSN) that is provided for all types of orders. Each separate order possesses a unique Order Code. In case of executions, matching orders are linked via a Match Code and a Trade Code. Match Codes are not provided for trades against hidden orders. The direction, price and volume of an order or trades are indicated under Bid/Ask, Price and AggSize columns, respectively. A "B" under the Bid/Ask column indicates a buy order while a "S" indicates a sell order. Trade volume and trade price for match orders are indicated under Trade Size and Trade Volume columns. The type of order is listed under the Market Mechanism Type (MMT) column. A “LO” represents a limit order while a “MO” stands for a market order.

The standing order book on the first day of the data is initialized via the “Broadcast Update Action” (BUA) parameter. A value of “F” in the BUA column for an Order Detail indicates previous orders at the start of the day that has not been fully matched, deleted or expired. For these orders, the order size equals the remaining size of the original trade submitted. Following the initialization process, subsequent orders are added or removed from the order book via the Order Detail and Order History files, respectively. There is no separate mechanism for altering normal orders; changes to an order are conveyed with two consecutive orders namely, a deletion followed by submission of a new order. Order Detail and Order History rows contain an “Order Action Type” parameter which is merged into the BUA column here for sake of brevity. Regular order submissions are realized via a value of “A” while Order Detail rows with a BUA of type “Z” constitute an exception where changes to volume information of a given order can be realized. Order History rows with BUA values of “D”, “E”, “P”, “M” and “T” represent deletions, expiries, partially filled orders, matched orders and transaction limits, respectively. Table A.1 provides a sample order matching sequence.

Table A.1: Sample Order Matching Table

Timestamp	MSN	DataType	OrderCode	MatchCode	TradeCode	Bid/Ask	Price	AggSize	TradeSize	TradePrice	MMT	BUA
6/1/2007 08:00:06	7209	OrderDetail	209UTBE107			S	934	285			LO	A
6/1/2007 08:00:06	7210	OrderDetail	309WJWD507			S	934.5	400			LO	A
6/1/2007 08:00:11	7312	OrderHistory	50ACLVAX07			B	932	65443	0		LO	D
6/1/2007 08:00:22	8095	OrderHistory	007K5Q2C07	60AMEO0B07	50ACNA7C07	S	0	0	5030	933	MO	M
6/1/2007 08:00:22	8096	OrderHistory	60AMEO0B07	007K5Q2C07	50ACNA7C07	B	0	6949	5030	933	MO	P
6/1/2007 08:00:22	8098	TradeReport			50ACNA7C07				5030	933		E
6/1/2007 08:00:22	8105	OrderHistory	309WK03D07	60AMEO0B07	50ACNA7F07	S	0	0	4073	933	MO	M
6/1/2007 08:00:22	8106	OrderHistory	60AMEO0B07	309WK03D07	50ACNA7F07	B	0	2876	4073	933	MO	P
6/1/2007 08:00:22	8108	TradeReport			50ACNA7F07				4073	933		E

Table A.1 shows that Trade Reports only arrive after the two corresponding order history rows have been created. All three rows, namely 2 Order History and 1 Trade Report, have matching codes on their Trade Code column. Occasionally more than a single transaction falls within a given second. In cases like these, to arrive at a fair measure of transaction price, a volume-weighted price is computed. The volume-weighted price for multiple transactions on a single second can be computed by:

$$\frac{\sum_{j}^{m} \sum P_j V_j}{\sum_{j}^{m} V_j}. \tag{A.1}$$

where m is the total number of transactions in a given second, P_j is the price of the j^{th} transaction and V_j is the volume of the j^{th} transaction.

Notice however, more than 3 rows have the same Trade Code in Table A.1. This presents us with an occasion where a single aggressor matches with multiple orders on the order book. In cases like these, the transaction is classified as a single trade with volume equal to the sum of all corresponding orders.

The sheer size of the raw data described above makes it impossible to load every data point into standard statistical packages. For this reason, a two step procedure is followed to sort and reorganize the SETS data. First, the raw data contained within the .csv files are loaded into the corresponding tables in MySQL database. This is an essential step in linking the information contained within each of the three tables and retrieving the necessary records for a given stock. Following the pooling of data, separate selection queries are produced for each of the 10 most liquid single stocks traded at the London Stock Exchange and a master table for each stock is produced containing information on order details, history and trades. The list of stocks used are presented in Appendix B.

In the second step, the master table for each selected stock is fed into Matlab, where orders added or removed from the orderbook according to their "action type". To make sure only orders present in the orderbook are removed from the dataset, orders are matched according to their unique order codes. Since the objective of this research is not to re-enact how the Exchange matches each submitted order but to have a snapshot of the whole orderbook at each time a trade occurs, a "bulk" approach is employed. To put it in another way, the code used to reconstruct the tables makes use of the matching trade codes within the framework of determining the initiator of a given trade rather than supply knowledge on which specific order would match another. Undertaking such a task would be redundant as the Exchange sorts orders according to their price and submission time and provides the order matching details in the trade report files.

Unlike the simply calculated values for price or trade volume, the determination of a trade initiator deserves some explanation. The initiator of a trade is determined with the following algorithm. In cases where a limit order matches with a market order, the direction of the market order is taken as the initiator. On the other hand, when two limit orders match, the initiator is designated to be the later arriving order as the other order has been standing in the orderbook. Similarly, for public orders that match hidden orders, the initiator is taken to be the public order. To remove multiple instances for the same order that has matched more than one order, trades which have the same initiator and bid/ask tag are joined as one effectively reducing the number of trades initiated while keeping volume information intact.

There are also cases when two market orders match each other, due to the opening auction. In such cases, the first arriving market order is taken as the initiator. However, since the opening auction is not part of the "normal" trading hours, initiator classification during these hours are of low importance as the dataset pertaining to the first 5 minutes following the commence of the regular trading hours is discarded. This is a necessary step to obtain reliable data free of the contamination during the opening interval.

In the end, a series of snapshots are produced at each trade (tick) time with information on the whole orderbook as well as traded quantity and volume weighted price. The time series data produced by the above two-step procedure now enables one to test the assumption of normality under the subordinator introduced in Equation (2.32).

Appendix B

Top 10 FTSE 100 Firms by Market Capitalization (As of June 18th 2012)

1. HSBC Holdings
2. Vodafone Group
3. BP
4. GlaxoSmithKline
5. British American Tobacco
6. Royal Dutch Shell
7. BG Group
8. Rio Tinto
9. Diageo
10. SAB Miller

Appendix C

Table C.1: Subordination Results for British American Tobacco Stock[28]

Linear	**Q1** (100 Ticks)	**Q2** (100 Ticks)	**Q3** (100 Ticks)	**Q4** (100 Ticks)
μ	-1.1521 e-7 (0)	2.5634 e-4 (0.0469)	-8.3825 e-8 (0)	- 2.3570 e-4 (0.1469)
σ	3.2309 e-7 (0.9990)	0.0040 (0)	3.3020 e-7 (0.9986)	0.0054 (0)
Log-Volume	-2.2954 e+19 (0)	9.9927 e+9 (0.0033)	-	-
Duration	2.0499 e+19 (0)	1.0000 e+10 (0)	2.2049 e+19 (0)	1.0000 e+10 (0)
Log-InitImb^2	4.6636 e+18 (0)	9.9905 e+9 (0)	-3.3199 e+18 (0)	9.8255 e+9 (0)
Log-VolImb^2	4.6636 e+18 (0)	-	7.0822 e+19 (0)	-
Log-likelihood	-10,987	-3,924	-18,782	-4,235

Autoregressive	**Q1** (100 Ticks)	**Q2** (100 Ticks)	**Q3** (100 Ticks)	**Q4** (100 Ticks)
μ	-7.0378 e-11 (0.0002)	1.4067 e-9 (0.3908)	4.2779 e-8 (0)	4.7413 e-9 (0.0018)
σ	5.3815 e-10 (1)	5.0543 e-8 (0.9998)	7.3008 e-8 (0.9997)	5.0818 e-8 (0.9998)
$r_{tick}^2(t-1)$	4.9529 e+19 (0)	4.5214 e+19 (0.0001)	8.4470 e+19 (0)	8.9796 e+19 (0)
Log-Volume	-	-	-	-
Duration	4.7589 e+19 (0)	8.1599 e+19 (0)	9.9942 e+19 (0)	9.0038 e+19 (0)
Log-InitImb^2	2.7338 e+19 (0)	5.8610 e+19 (0)	4.7465 e+19 (0)	7.9725 e+19 (0)
Log-VolImb^2	-	-	-	-
Log-likelihood	-15,578	-14,695	-20,179	-17,134

Normality		**Q1**		**Q2**		**Q3**		**Q4**	
		Linear	*Autoregressive*	*Linear*	*Autoregressive*	*Linear*	*Autoregressive*	*Linear*	*Autoregressive*
KS Test		0.0546 (0.0163)	0.0617 (0.0046)	0.0471 (0.0279)	0.0467 (0.0303)	0.0428 (0.0140)	0.0420 (0.0167)	0.0428 (0.0330)	0.0414 (0.0429)
JB Test		1,288 (0.0010)	232 (0.0010)	107 (0.0010)	154 (0.0010)	1,485 (0.0010)	327 (0.0010)	171 (0.0010)	44 (0.0010)
GARCH	KS Test	0.0513 (0.0301)		0.0429 (0.0577)		0.0407 (0.0223)		0.0408 (0.0476)	
	JB Test	39 (0.0010)		99 (0.0010)		75 (0.0010)		3 (0.2280)	

LB Test	**Q1**		**Q2**		**Q3**		**Q4**	
	Linear	*Autoregressive*	*Linear*	*Autoregressive*	*Linear*	*Autoregressive*	*Linear*	*Autoregressive*
R_{tick}	39.9191 (0.0051)	28.7350 (0.0931)	20.4126 (0.4324)	19.6020 (0.4831)	36.7815 (0.0124)	21.6975 (0.3571)	16.5722 (0.6806)	18.0809 (0.5821)
R_{tick}^2	148.6908 (0)	79.7184 (4.3825 e-9)	19.3393 (0.4999)	13.7170 (0.8445)	166.4169 (0)	110.3693 (1.6764 e-14)	69.0445 (2.6066 e-7)	23.9795 (0.2433)

[28] All values in parentheses throughout Appendix C show respective p-values for each variable.

Table C.2: Subordination Results for BG Group Stock

Linear	**Q1** (100 Ticks)	**Q2** (100 Ticks)	**Q3** (100 Ticks)	**Q4** (100 Ticks)
μ	5.2752 e-9 (0.4775)	1.2289 e-8 (0)	-1.7600 e-9 (0.5236)	-6.3670 e-8 (0)
σ	2.2900 e-7 (0.9992)	9.9620 e-8 (0.9996)	1.0177 e-7 (0.9996)	1.9840 e-7 (0.9992)
Log-Volume	-	-	9.7448 e+19 (0)	-
Duration	-6.6856 e+17 (0)	2.8412 e+19 (0)	5.9133 e+19 (0)	3.0306 e+19 (0)
Log-InitImb²	1.6388 e+18 (0)	2.0627 e+19 (0)	1.0000 e+20 (0)	6.9642 e+18 (0)
Log-VolImb²	2.2115 e+19 (0)	1.0000 e+20 (0)	-	-
Log-likelihood	-13,191	-16,022	-19,336	-16,280

Autoregressive	**Q1** (100 Ticks)	**Q2** (100 Ticks)	**Q3** (100 Ticks)	**Q4** (100 Ticks)
μ	-1.6080 e-8 (0)	3.7530 e-9 (0.0247)	-3.8989 e-9 (0.0294)	2.6111 e-10 (0.8797)
σ	5.1336 e-8 (0.9998)	5.5061 e-8 (0.9998)	6.4588 e-8 (0.9997)	5.8724 e-8 (0.9998)
$r_{tick}^2(t-1)$	9.7638 e+19 (0)	9.9365 e+19 (0)	8.5969 e+19 (0)	9.8878 e+19 (0)
Log-Volume	-	9.6220 e+19 (0)	9.8896 e+19 (0)	7.6725 e+19 (0)
Duration	6.9329 e+19 (0)	6.4660 e+19 (0)	8.7163 e+19 (0)	7.5697 e+19 (0)
Log-InitImb²	1.0000 e+20 (0)	9.5759 e+19 (0)	9.2383 e+19 (0)	8.6997 e+19 (0)
Log-VolImb²	-	-	-	-
Log-likelihood	-14,690	-16,571	-19,658	-17,533

Normality		**Q1**		**Q2**		**Q3**		**Q4**	
		Linear	*Autoregressive*	*Linear*	*Autoregressive*	*Linear*	*Autoregressive*	*Linear*	*Autoregressive*
KS Test		0.0476 (0.0261)	0.0469 (0.0296)	0.0638 (2.7226 e-4)	0.0527 (0.0046)	0.0731 (1.6497 e-6)	0.0706 (4.2399 e-6)	0.0520 (0.0037)	0.0499 (0.0059)
JB Test		88 (0.0010)	508 (0.0010)	5,712 (0.0010)	1,669 (0.0010)	17,057 (0.0010)	40,037 (0.0010)	635 (0.0010)	197 (0.0010)
GARCH	KS Test	0.0432 (0.0554)		0.0432 (0.0334)		0.0519 (0.0017)		0.0490 (0.0075)	
	JB Test	59 (0.0010)		752 (0.0010)		1,647 (0.0010)		38 (0.0010)	

LB Test	**Q1**		**Q2**		**Q3**		**Q4**	
	Linear	*Autoregressive*	*Linear*	*Autoregressive*	*Linear*	*Autoregressive*	*Linear*	*Autoregressive*
R_{tick}	44.1997 (0.0014)	33.6284 (0.0288)	21.1028 (0.3911)	18.1079 (0.5803)	40.2633 (0.0046)	27.7002 (0.1167)	21.5000 (0.3682)	21.0480 (0.3943)
R_{tick}^2	110.3667 (1.6875 e-14)	45.5594 (9.2608 e-4)	133.9519 (0)	148.7570 (0)	80.8810 (2.7808 e-9)	13.5016 (0.8548)	117.0909 (9.9920 e-16)	17.5531 (0.6168)

Table C.3: Subordination Results for British Petroleum Stock

Linear	**Q1** (100 Ticks)	**Q2** (100 Ticks)	**Q3** (100 Ticks)	**Q4** (100 Ticks)
μ	5.4774 e-10 (0.8499)	5.6039 e-9 (0.0001)	-1.1944 e-7 (0)	1.6878 e-9 (0.1992)
σ	1.0545 e-7 (0.9996)	5.2983 e-8 (0.9998)	1.5368 e-7 (0.9993)	4.6938 e-8 (0.9998)
Log-Volume	-	-	1.0000 e+20 (0.0003)	1.9225 e+19 (0.0193)
Duration	9.9999 e+19 (0)	3.3916 e+19 (0)	7.1360 e+19 (0.0080)	9.2349 e+19 (0)
Log-InitImb^2	6.5665 e+18 (0.0017)	8.3459 e+19 (0)	4.3670 e+19 (0.0002)	9.6643 e+19 (0)
Log-VolImb^2	-	3.7251 e+19 (0)	-	-
Log-likelihood	-19,881	-20,013	-23,179	-19,829

Autoregressive	**Q1** (100 Ticks)	**Q2** (100 Ticks)	**Q3** (100 Ticks)	**Q4** (100 Ticks)
μ	2.4504 e-10 (0.8214)	2.2236 e-8 (0)	1.0981 e-10 (0.9234)	-4.8625 e-9 (0)
σ	3.9512 e-8 (0.9998)	3.3979 e-8 (0.9999)	4.5813 e-8 (0.9998)	3.5850 e-8 (0.9998)
$r_{tick}^2(t-1)$	9.9984 e+19 (0)	9.2228 e+19 (0)	9.9255 e+19 (0)	9.4747 e+19 (0)
Log-Volume	9.9971 e+19 (0)	1.0000 e+20 (0.0001)	6.8373 e+19 (0)	7.6953 e+19 (0)
Duration	-1.1684 e+15 (0)	9.9660 e+19 (0)	9.5960 e+19 (0)	8.7638 e+19 (0)
Log-InitImb^2	9.9867 e+19 (0)	9.9314 e+19 (0)	9.9427 e+19 (0)	7.8021 e+19 (0)
Log-VolImb^2	-	-	-	-
Log-likelihood	-20,713	-20,354	-24,781	-20,033

Normality		**Q1**		**Q2**		**Q3**		**Q4**	
		Linear	*Autoregressive*	*Linear*	*Autoregressive*	*Linear*	*Autoregressive*	*Linear*	*Autoregressive*
KS Test		0.0771 (2.5298 e-7)	0.0895 (1.0209 e-9)	0.0843 (1.5746 e-8)	0.0807 (7.5652 e-8)	0.0605 (1.4373 e-5)	0.0629 (5.4956 e-6)	0.0763 (6.0209 e-7)	0.0728 (2.2870 e-6)
JB Test		825 (0.0010)	35 (0.0010)	327 (0.0010)	39 (0.0010)	227 (0.0010)	401 (0.0010)	1,390 (0.0010)	681 (0.0010)
GARCH	KS Test	0.0755 (4.8584 e-7)		0.0716 (2.8945 e-6)		0.0599 (1.8433 e-5)		0.0596 (2.1262 e-4)	
	JB Test	39 (0.0010)		55 (0.0010)		34 (0.0010)		51 (0.0010)	

LB Test	**Q1**		**Q2**		**Q3**		**Q4**	
	Linear	*Autoregressive*	*Linear*	*Autoregressive*	*Linear*	*Autoregressive*	*Linear*	*Autoregressive*
R_{tick}	38.3925 (0.0079)	22.3978 (0.3193)	25.2947 (0.1904)	21.6862 (0.3578)	42.4340 (0.0024)	37.8834 (0.0092)	26.0675 (0.1636)	26.7222 (0.1433)
R_{tick}^2	817 (0)	88.4018 (1.4088 e-10)	160.0943 (0)	48.8888 (3.1881 e-4)	331.9561 (0)	175.3197 (0)	12.1073 (0.9123)	13.3700 (0.8610)

Table C.4: Subordination Results for Diageo Stock

Linear	**Q1** (100 Ticks)	**Q2** (100 Ticks)	**Q3** (100 Ticks)	**Q4** (100 Ticks)
μ	2.0497 e-9 (0.2892)	-1.5261 e-14 (0.6246)	-1.1230 e-7 (0)	-3.9786 e-9 (0.1307)
σ	5.5268 e-8 (0.9998)	9.1326 e-13 (1)	1.8411 e-7 (0.9993)	8.3500 e-8 (0.9997)
Log-Volume	6.6499 e+19 (0)	-	-	-
Duration	4.5791 e+19 (0)	1.4486 e+29 (0)	5.0329 e+19 (0.0130)	1.6162 e+19 (0)
Log-InitImb²	5.6861 e+19 (0)	1.8021 e+29 (0)	6.8562 e+19 (0.0004)	-1.5207 e+18 (0)
Log-VolImb²	-	2.1857 e+29 (0.0003)	9.7174 e+19 (0.0007)	1.0000 e+20 (0)
Log-likelihood	-12,494	-22,570	-17,865	-15,023

Autoregressive	**Q1** (100 Ticks)	**Q2** (100 Ticks)	**Q3** (100 Ticks)	**Q4** (100 Ticks)
μ	-5.2023 e-10 (0.7626)	-4.4156 e-9 (0.0015)	1.0961 e-8 (0)	-1.8292 e-8 (0)
σ	4.9198 e-8 (0.9998)	4.0606 e-8 (0.9999)	5.3647 e-8 (0.9998)	4.1070 e-8 (0.9998)
$r_{tick}^2(t-1)$	7.5215 e+19 (0)	9.9986 e+19 (0)	7.2471 e+19 (0)	9.5429 e+19 (0)
Log-Volume	8.9904 e+19 (0)	-	-	-
Duration	3.1567 e+19 (0)	3.9005 e+19 (0)	9.9410 e+19 (0)	7.8746 e+19 (0)
Log-InitImb²	4.6724 e+19 (0)	9.5468 e+19 (0)	9.8314 e+19 (0)	8.9904 e+19 (0)
Log-VolImb²	-	8.4758 e+19 (0.0009)	6.1026 e+19 (0)	1.0000 e+20 (0.0008)
Log-likelihood	-12,545	-13,344	-19,045	-15,634

Normality		**Q1**		**Q2**		**Q3**		**Q4**	
		Linear	*Autoregressive*	*Linear*	*Autoregressive*	*Linear*	*Autoregressive*	*Linear*	*Autoregressive*
KS Test		0.0559 (0.0117)	0.0532 (0.0190)	0.0690 (5.3352 e-4)	0.0710 (3.3140 e-4)	0.0615 (1.5454 e-4)	0.0609 (1.8349 e-4)	0.0507 (0.0110)	0.0597 (0.0015)
JB Test		50 (0.0010)	34 (0.0010)	14 (0.0035)	5 (0.0640)	2,399 (0.0010)	3,432 (0.0010)	22 (0.0010)	8 (0.0196)
GARCH	KS Test	0.0503 (0.0308)		0.0695 (4.7165 e-4)		0.0575 (5.0766 e-4)		0.0490 (0.0154)	
	JB Test	6 (0.0421)		61 (0.0010)		122 (0.0010)		25 (0.0010)	

LB Test	**Q1**		**Q2**		**Q3**		**Q4**	
	Linear	*Autoregressive*	*Linear*	*Autoregressive*	*Linear*	*Autoregressive*	*Linear*	*Autoregressive*
R_{tick}	32.7175 (0.0362)	30.1502 (0.0675)	17.4250 (0.6252)	16.0873 (0.7112)	32.8498 (0.0350)	32.0940 (0.0423)	40.5332 (0.0043)	31.8546 (0.449)
R_{tick}^2	144.5198 (0)	44.9474 (0.0011)	50.5501 (1.8466 e-4)	30.5401 (0.0616)	314.6211 (0)	132.6829 (0)	56.0636 (2.8435 e-5)	21.9425 (0.3436)

Table C.5: Subordination Results for GlaxoSmith Kline Stock

Linear	**Q1** (100 Ticks)	**Q2** (100 Ticks)	**Q3** (100 Ticks)	**Q4** (100 Ticks)
μ	6.2051 e-8 (0)	5.8319 e-9 (0.0562)	9.1322 e-9 (0.0119)	3.0691 e-8 (0)
σ	2.4845 e-7 (0.9990)	1.0637 e-7 (0.9996)	1.4132 e-7 (0.9994)	1.9093 e-7 (0.9992)
Log-Volume	-	6.6333 e+18 (0)	-	1.0000 e+20 (0)
Duration	3.3822 e+19 (0)	3.9671 e+18 (0)	2.2637 e+19 (0.0156)	-9.1226 e+18 (0)
Log-InitImb^2	-1.0982 e+19 (0)	4.6393 e+19 (0)	4.2812 e+19 (0)	7.0501 e+19 (0)
Log-VolImb^2	9.9999 e+19 (0)	-	1.0000 e+20 (0)	-
Log-likelihood	-17,671	-17,770	-22,284	-17,300

Autoregressive	**Q1** (100 Ticks)	**Q2** (100 Ticks)	**Q3** (100 Ticks)	**Q4** (100 Ticks)
μ	-7.9494 e-9 (0)	1.5186 e-8 (0)	-2.7025 e-8 (0)	1.7543 e-8 (0)
σ	3.0683 e-8 (0.9999)	3.1238 e-8 (0.9999)	6.0855 e-8 (0.9997)	4.8753 e-8 (0.9998)
$r_{tick}^2(t-1)$	7.3744 e+19 (0)	6.0055 e+19 (0)	-	6.9436 e+19 (0)
Log-Volume	9.9945 e+19 (0)	8.8857 e+19 (0)	9.3870 e+19 (0)	-
Duration	9.5550 e+19 (0)	7.2474 e+19 (0)	1.7066 e+19 (0.0055)	9.5990 e+19 (0)
Log-InitImb^2	9.5550 e+19 (0)	9.9937 e+19 (0)	9.9181 e+19 (0)	5.4874 e+19 (0)
Log-VolImb^2	-	-	8.1451 e+19 (0)	9.8412 e+19 (0)
Log-likelihood	-19,344	-18,883	-22,888	-18,399

Normality		**Q1**		**Q2**		**Q3**		**Q4**	
		Linear	*Autoregressive*	*Linear*	*Autoregressive*	*Linear*	*Autoregressive*	*Linear*	*Autoregressive*
KS Test		0.0774 (6.2649 e-7)	0.0898 (3.4869 e-9)	0.0704 (1.0974 e-5)	0.0661 (4.6097 e-5)	0.0492 (0.0012)	0.0534 (3.3456 e-4)	0.0567 (8.7779 e-4)	0.0520 (0.0030)
JB Test		6,265 (0.0010)	6,604 (0.0010)	545 (0.0010)	65 (0.0010)	1,880 (0.0010)	352 (0.0010)	32 (0.0010)	61 (0.0010)
GARCH	KS Test	0.0809 (1.5988 e-7)		0.0573 (6.6848 e-4)		0.0422 (0.0087)		0.0464 (0.0113)	
	JB Test	7,384 (0.0010)		58 (0.0010)		3,580 (0.0010)		6 (0.0390)	

LB Test	**Q1**		**Q2**		**Q3**		**Q4**	
	Linear	*Autoregressive*	*Linear*	*Autoregressive*	*Linear*	*Autoregressive*	*Linear*	*Autoregressive*
R_{tick}	26.4687 (0.1509)	34.5167 (0.0228)	18.9435 (0.5255)	19.0396 (0.5193)	76.9710 (1.2748 e-8)	55.4953 (3.4616 e-5)	45.5234 (9.3661 e-4)	35.4982 (0.0176)
R_{tick}^2	38.1343 (0.0085)	27.4576 (0.1229)	26.0144 (0.1653)	32.8649 (0.0349)	323.4220 (0)	60.5702 (5.8101 e-6)	26.2401 (0.1580)	24.1626 (0.2354)

Table C.6: Subordination Results for HSBC Stock

Linear	**Q1** (100Ticks)	**Q2** (100Ticks)	**Q3** (100Ticks)	**Q4** (100Ticks)
μ	2.7296 e-9 (0.0512)	-2.0346 e-9 (0.1408)	1.0424 e-7 (0)	2.7508 e-9 (0.9998)
σ	5.3767 e-8 (0.9998)	5.7397 e-8 (0.9997)	2.6790 e-7 (0.9986)	4.5997 e-4 (0)
Log-Volume	2.6997 e+19 (0)	1.3925 e+19 (0.0032)	6.9045 e+19 (0)	1.0000 e+18 (0)
Duration	5.0572 e+18 (0.0086)	7.1242 e+19 (0)	-2.2287 e+19 (0)	-1.0000 e+18 (0)
Log-InitImb²	7.3548 e+19 (0)	4.8892 e+19 (0)	5.5295 e+19 (0.0003)	1.0000 e+18 (0)
Log-VolImb²	-	-	6.0521 e+19 (0)	1.0000 e+19 (0)
Log-likelihood	-22,627	-26,360	-31,413	-11,400

Autoregressive	**Q1** (100Ticks)	**Q2** (100Ticks)	**Q3** (100Ticks)	**Q4** (100Ticks)
μ	8.7329 e-9 (0)	-3.0180 e-9 (0.0008)	5.1429 e-11 (0.9601)	1.2593 e-10 (0.9024)
σ	2.6825 e-8 (0.9999)	3.7524 e-8 (0.9998)	4.8381 e-8 (0.9997)	4.2150 e-8 (0.9998)
$r_{tick}^2(t-1)$	8.5851 e+19 (0)	3.5343 e+19 (0.0001)	8.5057 e+19 (0)	6.1408 e+19 (0)
Log-Volume	9.5569 e+19 (0)	-	6.3423 e+19 (0.0235)	8.8872 e+19 (0)
Duration	1.0000 e+20 (0)	9.2838 e+19 (0)	5.2372 e+19 (0.0001)	-2.7778 e+19 (0)
Log-InitImb²	6.8892 e+19 (0)	9.8567 e+19 (0)	6.6997 e+19 (0)	1.4235 e+19 (0)
Log-VolImb²	-	7.8435 e+19 (0)	9.4366 e+19 (0.0006)	9.1453 e+19 (0)
Log-likelihood	-23,508	-27,069	-34,319	-26,224

Normality		**Q1**		**Q2**		**Q3**		**Q4**	
		Linear	*Autoregressive*	*Linear*	*Autoregressive*	*Linear*	*Autoregressive*	*Linear*	*Autoregressive*
KS Test		0.0748 (1.2026 e-7)	0.0663 (4.2692 e-6)	0.0474 (8.2207 e-4)	0.0481 (6.5517 e-4)	0.0503 (2.4969 e-5)	0.0477 (7.7809 e-5)	0.0456 (0.0018)	0.0448 (0.0022)
JB Test		77 (0.0010)	139 (0.0010)	642 (0.0010)	238 (0.0010)	554 (0.0010)	1,248 (0.0010)	2 (0.2947)	0 (0.5000)
GARCH	KS Test	0.0529 (4.9487 e-4)		0.0539 (8.3033 e-5)		0.0395 (0.0019)		0.0415 (0.0059)	
	JB Test	6 (0.0430)		59 (0.0010)		57 (0.0010)		2 (0.4278)	

LB Test	**Q1**		**Q2**		**Q3**		**Q4**	
	Linear	*Autoregressive*	*Linear*	*Autoregressive*	*Linear*	*Autoregressive*	*Linear*	*Autoregressive*
R_{tick}	23.9789 (0.2433)	17.0894 (0.6472)	38.5263 (0.0076)	34.2855 (0.0243)	36.2941 (0.0142)	38.1251 (0.0085)	16.9765 (0.6545)	16.0319 (0.7146)
R_{tick}^2	175.8021 (0)	350.0387 (0)	216.8248 (0)	186.4615 (0)	271.0396 (0)	132.8735 (0)	30.2203 (0.0664)	21.9554 (0.3429)

Table C.7: Subordination Results for Rio Tinto Stock

Linear	**Q1** (100 Ticks)	**Q2** (100 Ticks)	**Q3** (100 Ticks)	**Q4** (100 Ticks)
μ	1.1684 e-9 (0.5711)	-9.4706 e-8 (0)	8.4961 e-8 (0)	-1.6292 e-7 (0)
σ	9.3839 e-8 (0.9995)	2.5467 e-7 (0.9986)	3.2634 e-7 (0.9981)	3.1648 e-7 (0.9982)
Log-Volume	1.1544 e+19 (0)	-	9.0359 e+19 (0)	9.5013 e+19 (0.0002)
Duration	-5.7860 e+17 (0)	9.9996 e+19 (0)	-	-2.0675 e+19 (0)
Log-InitImb²	6.1705 e+19 (0)	1.95138 e+19 (0)	9.7974 e+18 (0)	4.5420 e+19 (0.0126)
Log-VolImb²	1.0000 e+20 (0)	2.9427 e+19 (0)	-1.2440 e+18 (0)	9.7235 e+19 (0)
Log-likelihood	-30,559	-34,427	-38,358	-35,909

Autoregressive	**Q1** (100 Ticks)	**Q2** (100 Ticks)	**Q3** (100 Ticks)	**Q4** (100 Ticks)
μ	-1.0966 e-8 (0)	2.2256 e-9 (0.3529)	-8.2373 e-10 (0.5461)	5.3539 e-10 (0.6243)
σ	3.7744 e-8 (0.9998)	1.1858 e-7 (0.9993)	7.1668 e-8 (0.9996)	5.5568 e-8 (0.9997)
$r^2_{tick}(t-1)$	8.8421 e+19 (0)	4.2534 e+19 (0)	5.1620 e+19 (0)	7.3631 e+19 (0)
Log-Volume	9.9780 e+19 (0.0003)	-	6.5510 e+19 (0.0055)	7.3902 e+19 (0.0007)
Duration	8.3914 e+19 (0)	4.4047 e+19 (0)	7.7802 e+19 (0)	7.8727 e+19 (0)
Log-InitImb²	9.6529 e+19 (0)	4.5245 e+19 (0)	4.8523 e+19 (0)	4.5551 e+19 (0)
Log-VolImb²	9.9899 e+19 (0.0005)	4.4572 e+19 (0)	5.8612 e+19 (0.0196)	9.9918 e+19 (0)
Log-likelihood	-31,380	-35,522	-41,454	-39,514

Normality		**Q1**		**Q2**		**Q3**		**Q4**	
		Linear	*Autoregressive*	*Linear*	*Autoregressive*	*Linear*	*Autoregressive*	*Linear*	*Autoregressive*
KS Test		0.0371 (0.0065)	0.0412 (0.0017)	0.1154 (7.3105 e-29)	0.1186 (1.8939 e-30)	0.0558 (6.6182 e-8)	0.0472 (8.7838 e-6)	0.0341 (0.0048)	0.0355 (0.0029)
JB Test		667 (0.0010)	524 (0.0010)	5,669,668 (0.0010)	19,059,097 (0.0010)	4,692 (0.0010)	4,779 (0.0010)	234 (0.0010)	210 (0.0010)
GARCH	KS Test	0.0327 (0.0234)		0.0689 (1.4922 e-10)		0.0397 (3.2037 e-4)		0.0311 (0.0130)	
	JB Test	252 (0.0010)		61,557 (0.0010)		6,217 (0.0010)		159 (0.0010)	

LB Test	**Q1**		**Q2**		**Q3**		**Q4**	
	Linear	*Autoregressive*	*Linear*	*Autoregressive*	*Linear*	*Autoregressive*	*Linear*	*Autoregressive*
R_{tick}	34.2776 (0.0243)	27.4318 (0.1235)	116.4268 (1.3323 e-15)	103.0579 (3.5583 e-13)	38.1285 (0.0085)	27.0907 (0.1327)	19.4702 (0.4915)	14.3619 (0.8117)
R^2_{tick}	565.4356 (0)	354.4957 (0)	19.2040 (0.5086)	4.2789 (0.9999)	582.9675 (0)	279.0021 (0)	83.6051 (9.5141 e-10)	46.7978 (6.2571 e-4)

Table C.8: Subordination Results for SAB Miller Stock[29]

Linear	**Q1** (100 Ticks)	**Q2** (90 Ticks)	**Q3** (100 Ticks)	**Q4** (70 Ticks)
μ	8.9473 e-7 (0.2358)	-8.6463 e-8 (0)	6.6771 e-8 (0)	2.9076 e-8 (0)
σ	1.9255 e-5 (0.9776)	3.2818 e-7 (0.9989)	1.4066 e-7 (0.9995)	1.8632 e-7 (0.9993)
Log-Volume	9.9999 e+14 (0)	6.5458 e+19 (0)	1.0000 e+20 (0)	-
Duration	4.6693 e+14 (0)	-2.1460 e+18 (0)	1.1837 e+19 (0.0001)	5.7186 e+18 (0.0002)
Log-InitImb^2	1.0000 e+15 (0)	1.0000 e+20 (0)	2.7082 e+19 (0)	-3.9134 e+17 (0)
Log-VolImb^2	-	-	-1.2608 e+19 (0)	8.2618 e+19 (0)
Log-likelihood	-6,176	-12,649	-15,163	-17,707

Autoregressive	**Q1** (100 Ticks)	**Q2** (90 Ticks)	**Q3** (100 Ticks)	**Q4** (70 Ticks)
μ	1.5232 e-9 (0.4161)	1.5490 e-9 (0.4065)	5.9718 e-9 (0.0147)	-5.5842 e-9 (0.0005)
σ	4.7791 e-8 (0.9999)	5.6189 e-8 (0.9998)	7.9497 e-8 (0.9997)	5.5972 e-8 (0.9998)
$r_{tick}^2(t-1)$	6.7977 e+19 (0)	1.8335 e+19 (0.0029)	8.1994 e+19 (0)	1.0000 e+20 (0)
Log-Volume	-	9.9925 e+19 (0.0001)	9.9877 e+19 (0)	-
Duration	9.1697 e+19 (0)	6.8442 e+19 (0)	4.6871e+19 (0)	1.0000 e+20 (0)
Log-InitImb^2	9.9065 e+19 (0)	1.0000 e+20 (0)	2.8841 e+19 (0)	-3.4439 e+11 (0)
Log-VolImb^2	-	-	-	1.0000 e+20 (0)
Log-likelihood	-10,026	-13,865	-15,747	-18,791

Normality		**Q1**		**Q2**		**Q3**		**Q4**	
		Linear	*Autoregressive*	*Linear*	*Autoregressive*	*Linear*	*Autoregressive*	*Linear*	*Autoregressive*
KS Test		0.0453 (0.1318)	0.0482 (0.0929)	0.0545 (0.0087)	0.0438 (0.0597)	0.0397 (0.0699)	0.0393 (0.0745)	0.0506 (0.0035)	0.0510 (0.0032)
JB Test		85 (0.0010)	26 (0.0010)	340 (0.0010)	234 (0.0010)	129 (0.0001)	383 (0.0010)	25 (0.0010)	170 (0.0010)
GARCH	KS Test	0.0561 (0.0312)		0.0625 (0.0016)		0.0324 (0.2132)		0.0501 (0.0040)	
	JB Test	7 (0.0292)		572 (0.0010)		30 (0.0010)		43 (0.0010)	

LB Test	**Q1**		**Q2**		**Q3**		**Q4**	
	Linear	*Autoregressive*	*Linear*	*Autoregressive*	*Linear*	*Autoregressive*	*Linear*	*Autoregressive*
R_{tick}	19.3210 (0.5011)	21.0630 (0.3934)	18.7713 (0.5367)	23.9882 (0.2429)	25.1293 (0.1965)	21.7390 (0.3548)	26.2966 (0.1562)	27.1843 (0.1302)
R_{tick}^2	108.5900 (3.5527 e-14)	85.2687 (4.9200 e-10)	20.6774 (0.4163)	33.2963 (0.0313)	179.0044 (0)	183.8922 (0)	133.6843 (0)	47.6079 (4.8272 e-4)

[29] Different sampling frequencies have been used in Q2 and Q4 for SAB Miller as sparser sampling resulted in normally distributed returns without the need for subordination.

Table C.9: Subordination Results for Shell Stock

Linear	**Q1** (70 Ticks)	**Q2** (100 Ticks)	**Q3** (100 Ticks)	**Q4** (100 Ticks)
μ	4.4576 e-10 (0.9316)	-1.0791 e-10 (0.9349)	-3.7920 e-9 (0.0625)	8.5731 e-8 (0)
σ	1.9568 e-7 (0.9992)	4.2722 e-8 (0.9998)	7.5872 e-8 (0.9997)	1.9654 e-7 (0.9993)
Log-Volume	2.6920 e+19 (0)	1.0000 e+20 (0)	8.3493 e+19 (0)	6.5100 e+19 (0)
Duration	1.0000 e+20 (0)	1.0000 e+20 (0)	2.3232 e+19 (0)	-1.0927 e+18 (0)
Log-InitImb²	7.1329 e+19 (0.0001)	2.9118 e+19 (0)	9.8443 e+19 (0)	2.6871 e+19 (0)
Log-VolImb²	-4.0737 e+19 (0)	-	-2.3712 e+19 (0)	-
Log-likelihood	-20,550	-16,249	-20,801	-16,219

Autoregressive	**Q1** (70 Ticks)	**Q2** (100 Ticks)	**Q3** (100 Ticks)	**Q4** (100 Ticks)
μ	-6.1471 e-09 (0)	6.1171 e-8 (0)	-7.3507 e-10 (0.6480)	-4.5436 e-9 (0.0003)
σ	3.5711 e-08 (0.9998)	7.58291 e-8 (0.9997)	6.0001 e-8 (0.9997)	4.1984 e-8 (0.9998)
$r_{tick}^2(t-1)$	1.0000 e+20 (0)	6.5455 e+19 (0.0072)	1.2339 e+19 (0.0002)	5.2155 e+19 (0)
Log-Volume	7.0068 e+19 (0)	1.0000 e+20 (0)	4.8654 e+19 (0.0162)	9.9192 e+19 (0)
Duration	9.2564 e+19 (0)	7.9835 e+19 (0)	4.2889 e+19 (0)	7.1199 e+19 (0)
Log-InitImb²	8.2570 e+19 (0)	-2.2156 e+17 (0)	7.1681 e+19 (0)	7.9620 e+19 (0)
Log-VolImb²	-	-	6.5802 e+19 (0.0033)	-
Log-likelihood	-22,168	-15,635	-21,063	-17,512

Normality		**Q1**		**Q2**		**Q3**		**Q4**	
		Linear	*Autoregressive*	*Linear*	*Autoregressive*	*Linear*	*Autoregressive*	*Linear*	*Autoregressive*
KS Test		0.0459 (0.0049)	0.0462 (0.0045)	0.0417 (0.0513)	0.0017 (0.0510)	0.0383 (0.0331)	0.0307 (0.1424)	0.0379 (0.0758)	0.0379 (0.0756)
JB Test		2,282 (0.0010)	392 (0.0010)	26 (0.0010)	60 (0.0010)	525 (0.0010)	700 (0.0010)	22 (0.0001)	43 (0.0010)
GARCH	KS Test	0.0468 (0.0039)		0.0443 (0.0320)		0.0350 (0.0652)		0.0433 (0.0280)	
	JB Test	16 (0.0015)		32 (0.0010)		164 (0.0010)		17 (0.0013)	

LB Test	**Q1**		**Q2**		**Q3**		**Q4**	
	Linear	*Autoregressive*	*Linear*	*Autoregressive*	*Linear*	*Autoregressive*	*Linear*	*Autoregressive*
R_{tick}	75.3423 (2.3794 e-8)	56.0454 (2.8615 e-5)	17.0768 (0.6480)	17.2084 (0.6394)	21.9221 (0.3448)	20.7530 (0.4118)	13.2405 (0.8668)	12.9193 (0.8808)
R_{tick}^2	475.5357 (0)	260.4708 (0)	40.2501 (0.0046)	33.2846 (0.0314)	239.8774 (0)	145.2595 (0)	49.2733 (2.8118 e-4)	31.9224 (0.0441)

Table C.10: Subordination Results for Vodafone Stock

Linear	**Q1** (100 Ticks)	**Q2** (100 Ticks)	**Q3** (100 Ticks)	**Q4** (100 Ticks)
μ	8.4457 e-9 (0)	-6.0663 e-9 (0)	-1.2443 e-7 (0)	1.1147 e-6 (0.0003)
σ	5.1454 e-8 (0.9998)	5.5362 e-8 (0.9997)	2.1666 e-7 (0.9989)	1.3173 e-5 (0.9499)
Log-Volume	-2.5329 e+18 (0)	-2.5290 e+19 (0)	-2.1256 e+19 (0)	9.4929 e+14 (0.0385)
Duration	1.0000 e+20 (0)	1.4859 e+19 (0)	1.9632 e+19 (0)	8.1095 e+14 (0)
Log-InitImb^2	8.8501 e+19 (0)	7.1246 e+19 (0)	1.1806 e+19 (0.0154)	8.7256 e+14 (0)
Log-VolImb^2	3.5517 e+19 (0)	7.3204 e+19 (0)	8.1513 e+19 (0)	9.7458 e+14 (0.0218)
Log-likelihood	-23,605	-28,699	-31,011	-17,843

Autoregressive	**Q1** (100 Ticks)	**Q2** (100 Ticks)	**Q3** (100 Ticks)	**Q4** (100 Ticks)
μ	-1.4271 e-9 (0.1619)	-5.0652 e-9 (0)	-7.6819 e-11 (0)	-1.4458 e-9 (0.3502)
σ	3.9922 e-8 (0.9998)	3.0244 e-8 (0.9998)	1.3368 e-10 (1)	6.5692 e-8 (0.9997)
$r^2_{tick}(t-1)$	5.9664 e+19 (0)	8.7926 e+19 (0)	9.9675 e+24 (0)	4.2369 e+19 (0)
Log-Volume	-	9.1280 e+19 (0)	-	-
Duration	8.2030 e+19 (0)	9.5556 e+19 (0)	7.9695 e+24 (0)	4.6414 e+19 (0)
Log-InitImb^2	8.8930 e+19 (0)	4.3409 e+19 (0)	8.4025 e+24 (0)	1.3794 e+19 (0.0022)
Log-VolImb^2	7.5715 e+19 (0)	-	8.3709 e+24 (0)	4.2489 e+19 (0)
Log-likelihood	-23,858	-29,190	-45,945	-27,225

Normality		**Q1**		**Q2**		**Q3**		**Q4**	
		Linear	*Autoregressive*	*Linear*	*Autoregressive*	*Linear*	*Autoregressive*	*Linear*	*Autoregressive*
KS Test		0.0508 (7.1175 e-4)	0.0504 (7.9409 e-4)	0.0626 (7.5934 e-7)	0.0615 (1.2727 e-6)	0.0486 (6.3059 e-5)	0.0532 (8.1188 e-6)	0.0309 (0.0627)	0.0317 (0.0519)
JB Test		788 (0.0010)	380 (0.0010)	192 (0.0010)	125 (0.0010)	272 (0.0010)	242 (0.0010)	51 (0.0010)	82 (0.0010)
GARCH	KS Test	0.0453 (0.0036)		0.0572 (8.9762 e-6)		0.0456 (2.1554 e-4)		0.0341 (0.0294)	
	JB Test	115 (0.0010)		162 (0.0010)		133 (0.0010)		26 (0.0010)	

LB Test	**Q1**		**Q2**		**Q3**		**Q4**	
	Linear	*Autoregressive*	*Linear*	*Autoregressive*	*Linear*	*Autoregressive*	*Linear*	*Autoregressive*
R_{tick}	36.9369 (0.0119)	31.2321 (0.0522)	16.7988 (0.6660)	16.2786 (0.6992)	34.1910 (0.0249)	30.5675 (0.0612)	20.3761 (0.4346)	16.8092 (0.6653)
R^2_{tick}	296.7440 (0)	131.8968 (0)	37.6809 (0.0097)	25.8647 (0.1703)	427.1796 (0)	121.6263 (1.1102 e-16)	37.3605 (0.0106)	39.2457 (0.0062)

Appendix D

Table D.1.1: In-Sample Linear Crash Estimator LDA Confusion Matrices for E-Mini S&P500 Futures

Crash Threshold: -0.25%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	30	17	47
No Crash	149	583	732
Total	179	600	779

Crash Threshold: -0.50%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	9	0	9
No Crash	84	686	770
Total	93	686	779

Crash Threshold: -0.75%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	3	0	3
No Crash	90	686	776
Total	93	686	779

Crash Threshold: -1.00%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	2	0	2
No Crash	78	699	777
Total	80	699	779

Table D.1.2: Out-of-Sample Linear Crash Estimator LDA Confusion Matrices for E-Mini S&P500 Futures

Crash Threshold: -0.25%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	26	33	59
No Crash	226	495	721
Total	252	528	780

Crash Threshold: -0.50%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	2	4	6
No Crash	105	669	774
Total	107	673	780

Crash Threshold: -0.75%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	0	1	1
No Crash	50	729	779
Total	50	730	780

Crash Threshold: -1.00%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	0	0	0
No Crash	18	762	780
Total	18	762	780

Table D.2.1: In-Sample VPIN LDA Confusion Matrices for E-Mini S&P500 Futures

Crash Threshold: -0.25%

Actual		Predicted: Crash	Predicted: No Crash	Total
	Crash	33	14	47
	No Crash	332	400	732
	Total	365	414	779

Crash Threshold: -0.50%

Actual		Predicted: Crash	Predicted: No Crash	Total
	Crash	6	3	9
	No Crash	371	399	770
	Total	377	402	779

Crash Threshold: -0.75%

Actual		Predicted: Crash	Predicted: No Crash	Total
	Crash	2	1	3
	No Crash	356	420	776
	Total	358	421	779

Crash Threshold: -1.00%

Actual		Predicted: Crash	Predicted: No Crash	Total
	Crash	2	0	2
	No Crash	399	378	777
	Total	401	378	779

Table D.2.2: Out-of-Sample VPIN LDA Confusion Matrices for E-Mini S&P500 Futures

Crash Threshold: -0.25%

Actual		Predicted: Crash	Predicted: No Crash	Total
	Crash	33	26	59
	No Crash	370	351	721
	Total	403	377	780

Crash Threshold: -0.50%

Actual		Predicted: Crash	Predicted: No Crash	Total
	Crash	4	2	6
	No Crash	354	420	774
	Total	358	422	780

Crash Threshold: -0.75%

Actual		Predicted: Crash	Predicted: No Crash	Total
	Crash	0	1	1
	No Crash	163	616	779
	Total	163	617	780

Crash Threshold: -1.00%

Actual		Predicted: Crash	Predicted: No Crash	Total
	Crash	0	0	0
	No Crash	151	629	780
	Total	151	629	780

Table D.3.1: In-Sample Market Heat LDA Confusion Matrices for E-Mini S&P500 Futures

Crash Threshold: -0.25%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	6	41	47
No Crash	13	719	732
Total	19	760	779

Crash Threshold: -0.50%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	2	7	9
No Crash	6	764	770
Total	8	771	779

Crash Threshold: -0.75%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	1	2	3
No Crash	18	758	776
Total	19	760	779

Crash Threshold: -1.00%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	1	1	2
No Crash	13	764	777
Total	14	765	779

Table D.3.2: Out-of-Sample Market Heat LDA Confusion Matrices for E-Mini S&P500 Futures

Crash Threshold: -0.25%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	7	52	59
No Crash	173	548	721
Total	180	600	780

Crash Threshold: -0.50%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	1	5	6
No Crash	90	684	774
Total	91	689	780

Crash Threshold: -0.75%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	0	1	1
No Crash	29	750	779
Total	29	751	780

Crash Threshold: -1.00%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	0	0	0
No Crash	0	780	780
Total	0	780	780

Table D.4.1: In-Sample Market Heat LDA Confusion Matrices for E-Mini S&P500 Futures Using Trade Duration

Crash Threshold: -0.25%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	16	31	47
No Crash	117	615	732
Total	133	646	779

Crash Threshold: -0.50%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	4	5	9
No Crash	18	752	770
Total	22	757	779

Crash Threshold: -0.75%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	1	2	3
No Crash	48	728	776
Total	49	730	779

Crash Threshold: -1.00%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	1	1	2
No Crash	20	757	777
Total	21	758	779

Table D.4.2: Out-of-Sample Market Heat LDA Confusion Matrices for E-Mini S&P500 Futures Using Trade Duration

Crash Threshold: -0.25%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	17	42	59
No Crash	255	466	721
Total	272	508	780

Crash Threshold: -0.50%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	2	4	6
No Crash	121	653	774
Total	123	657	780

Crash Threshold: -0.75%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	0	1	1
No Crash	30	749	779
Total	30	750	780

Crash Threshold: -1.00%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	0	0	0
No Crash	4	776	780
Total	4	776	780

Figure E.1.1: Classification Accuracy for LSE Stocks (Crash Threshold: -0.10%)

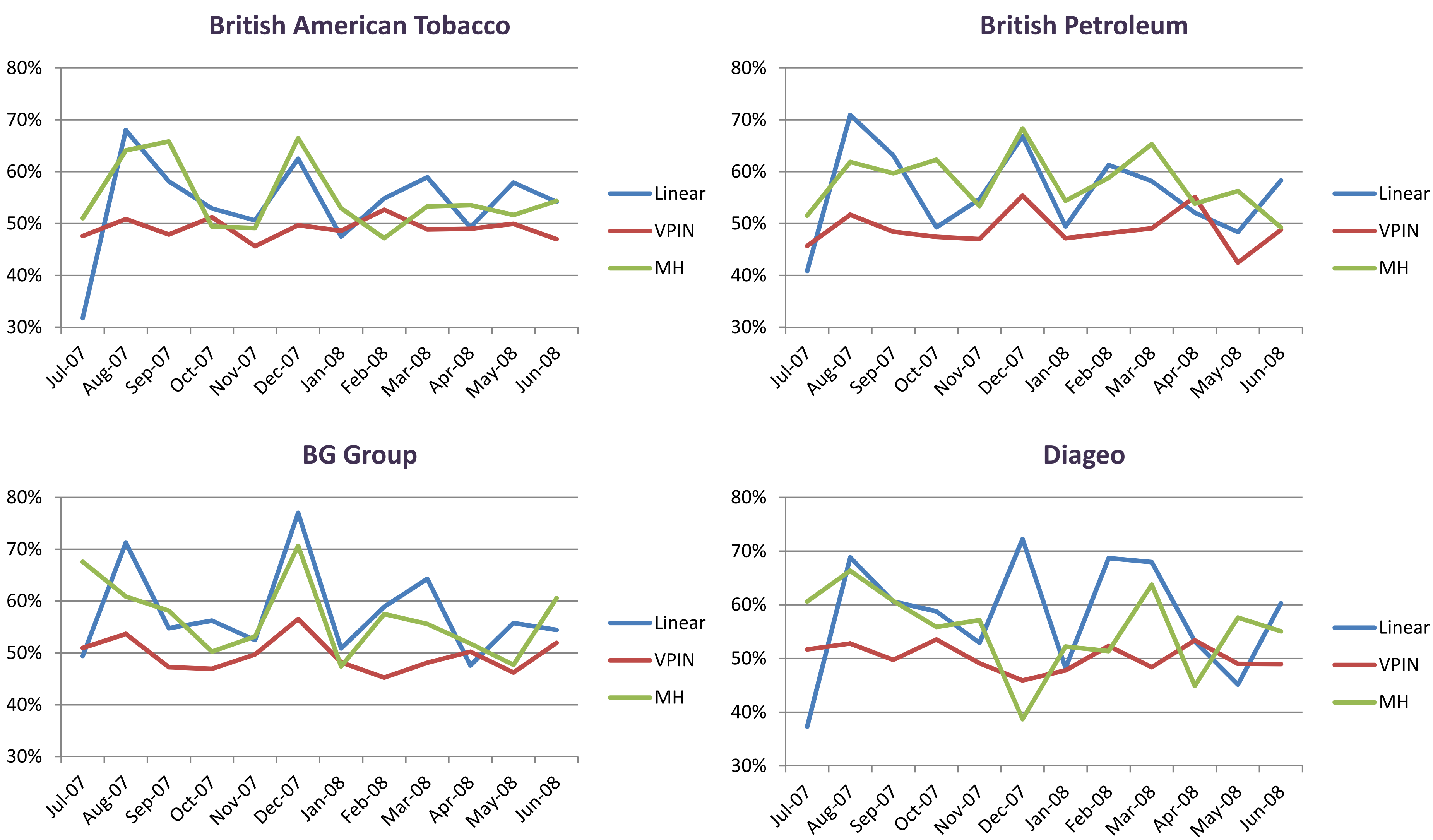


Figure E.1.2: Classification Accuracy for LSE Stocks (Crash Threshold: -0.10%)

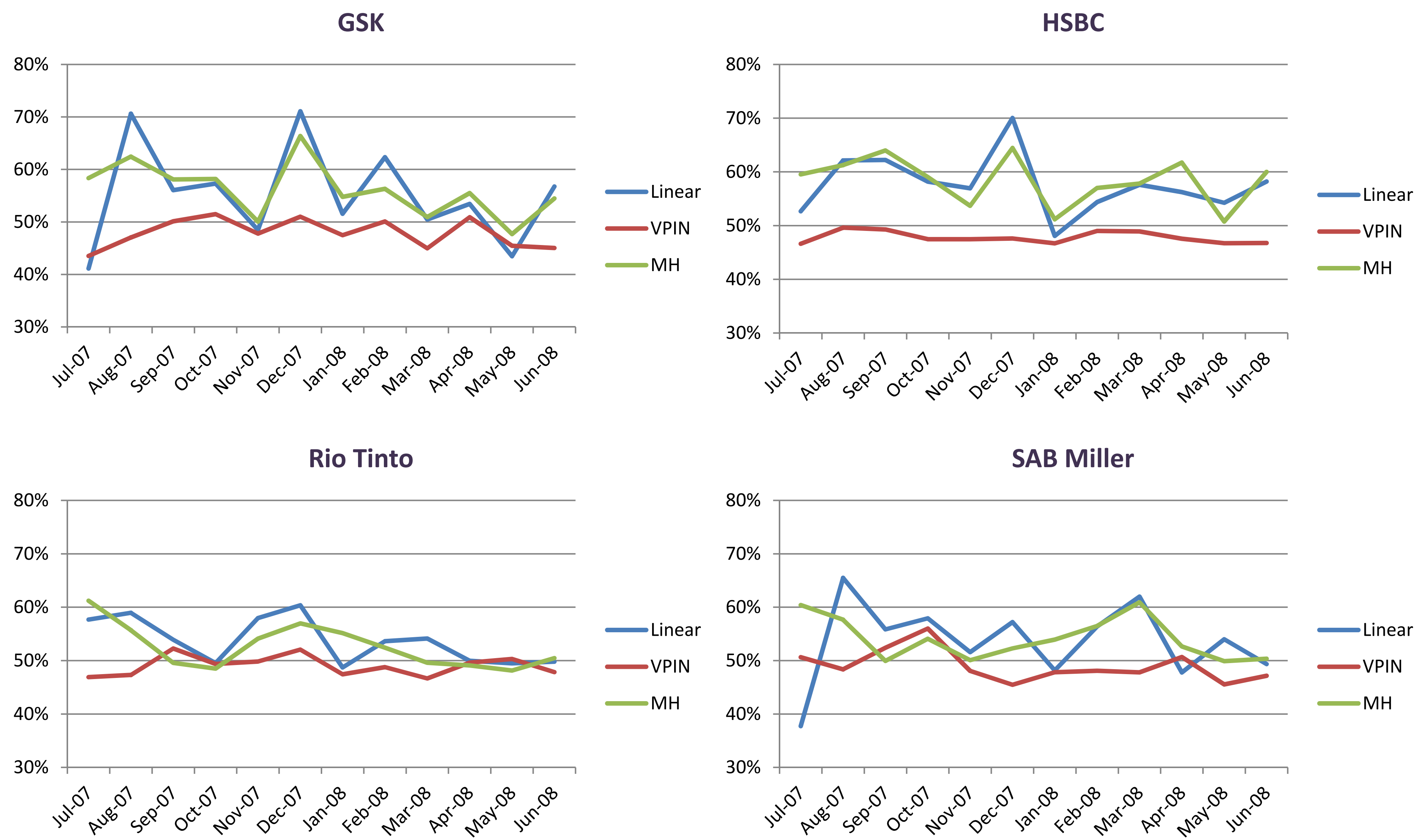


Figure E.1.3: Classification Accuracy for LSE Stocks (Crash Threshold: -0.10%)

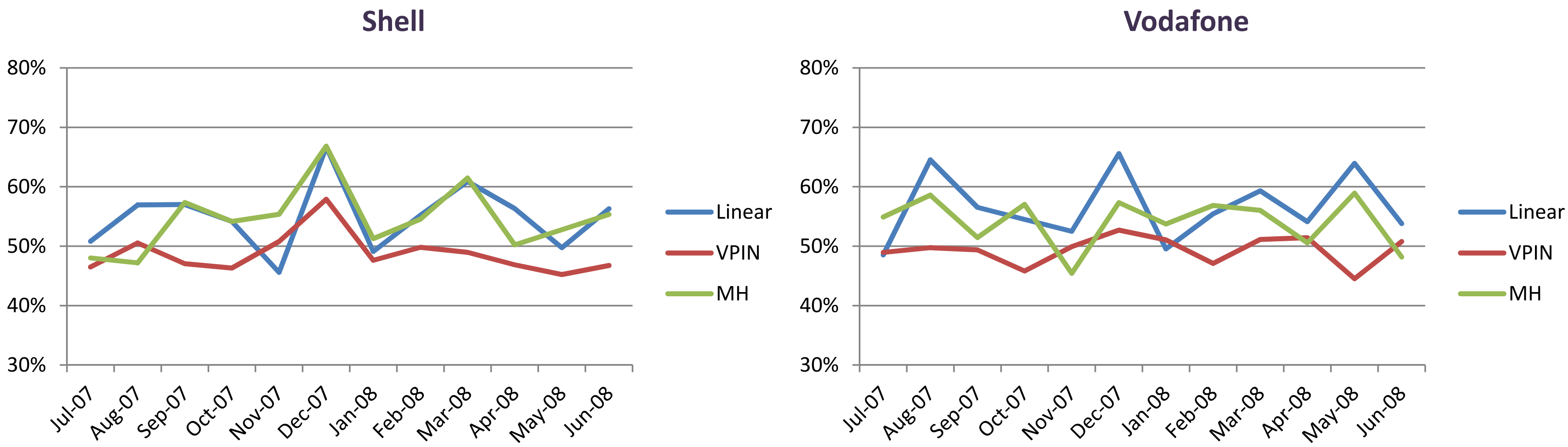


Figure E.2.1: Classification Accuracy for LSE Stocks (Crash Threshold: -0.50%)

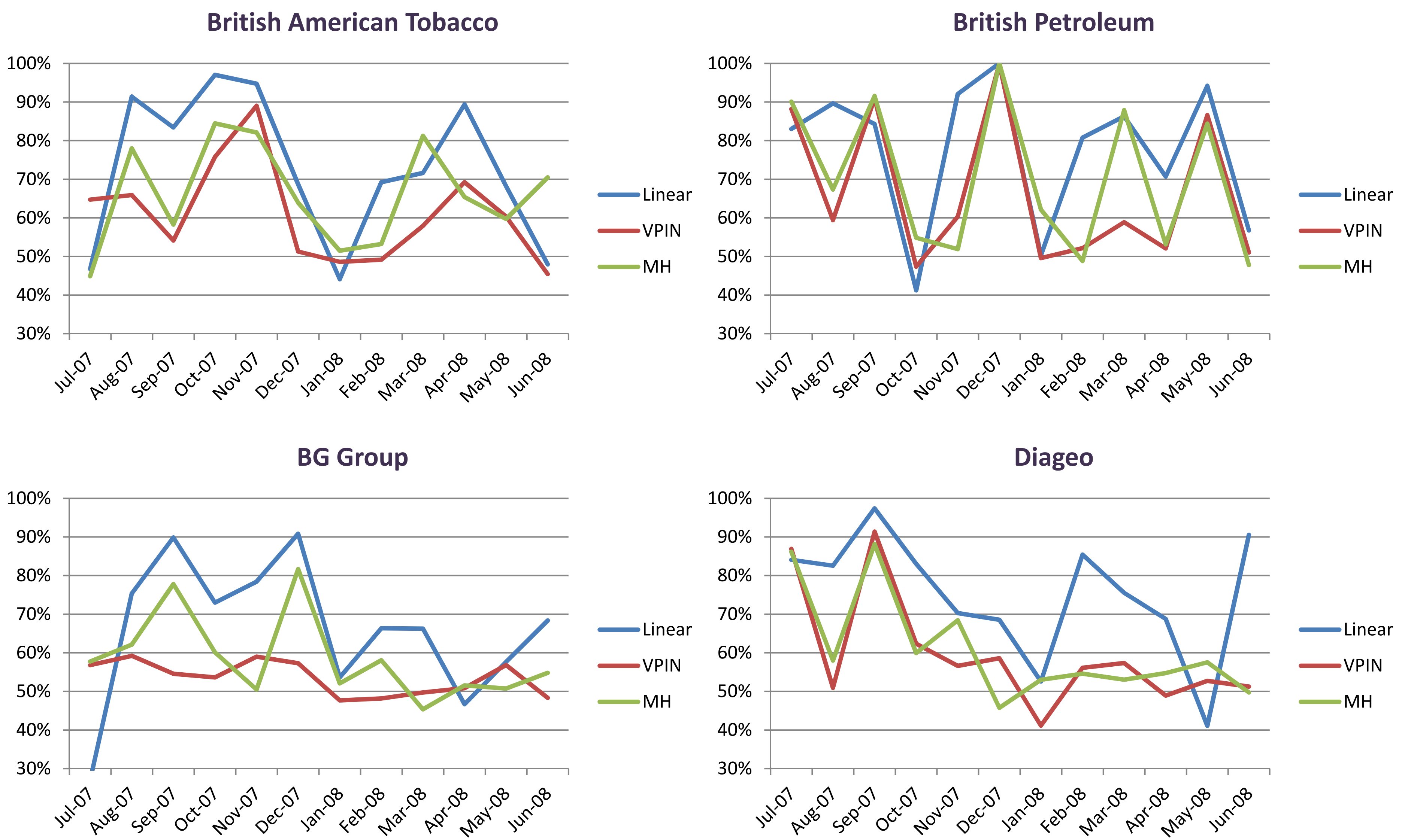


Figure E.2.2: Classification Accuracy for LSE Stocks (Crash Threshold: -0.50%)

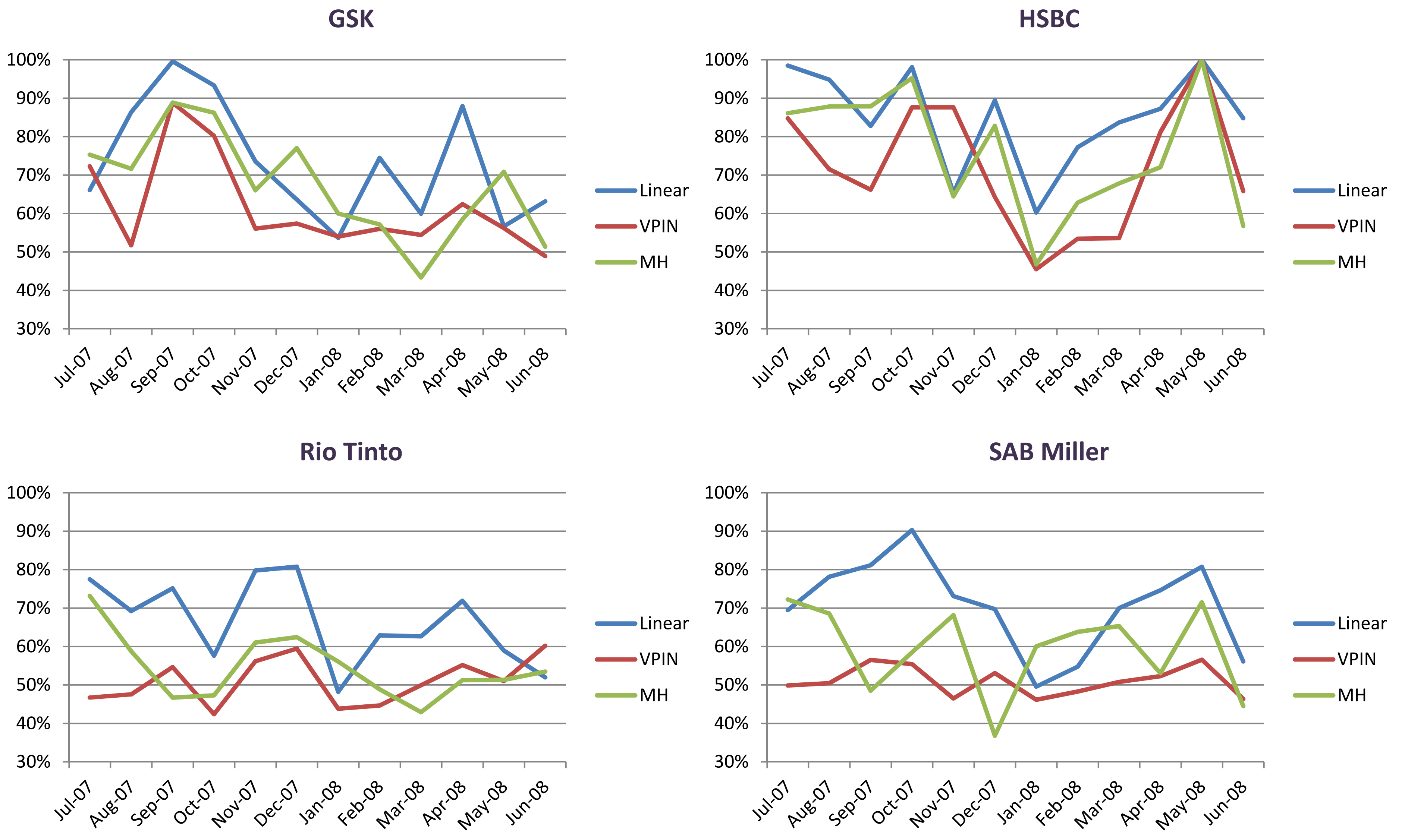
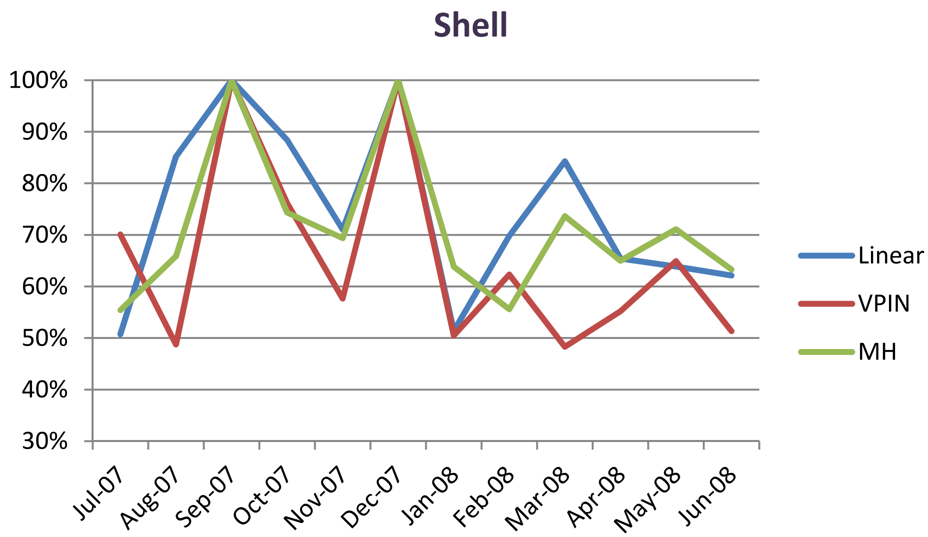
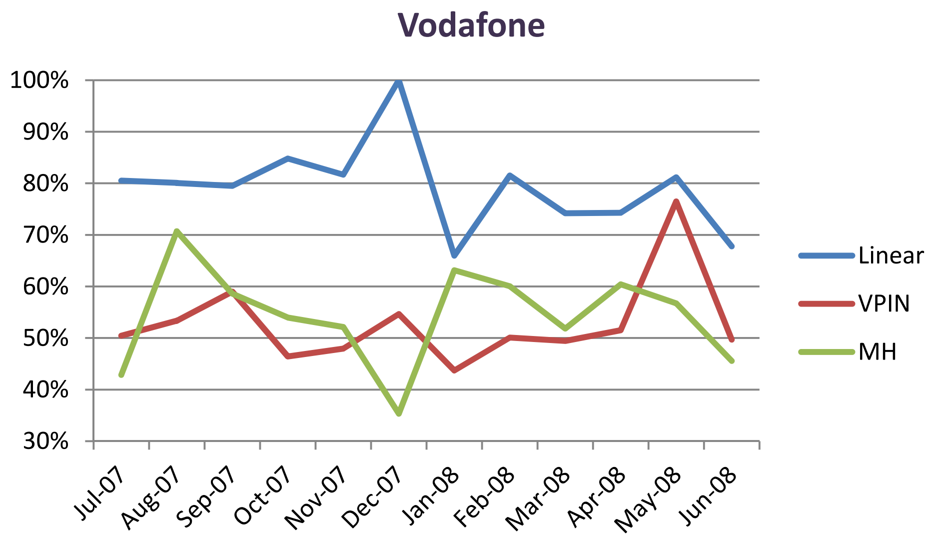


Figure E.2.3: Classification Accuracy for LSE Stocks (Crash Threshold: -0.50%)





Appendix F

Figure F.1.1: Precision for LSE Stocks (Crash Threshold: -0.10%)

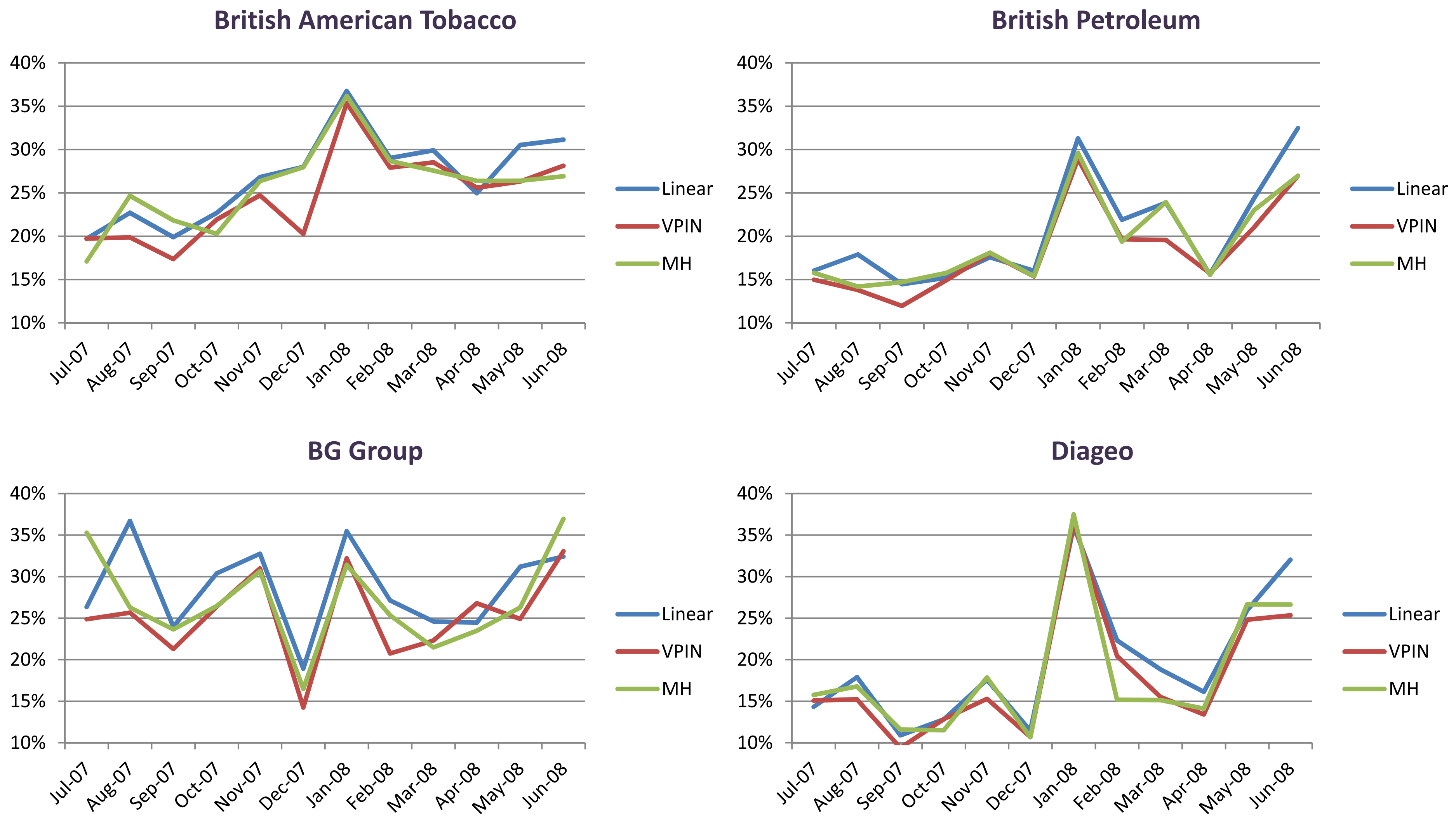


Figure F.1.2: Precision for LSE Stocks (Crash Threshold: -0.10%)

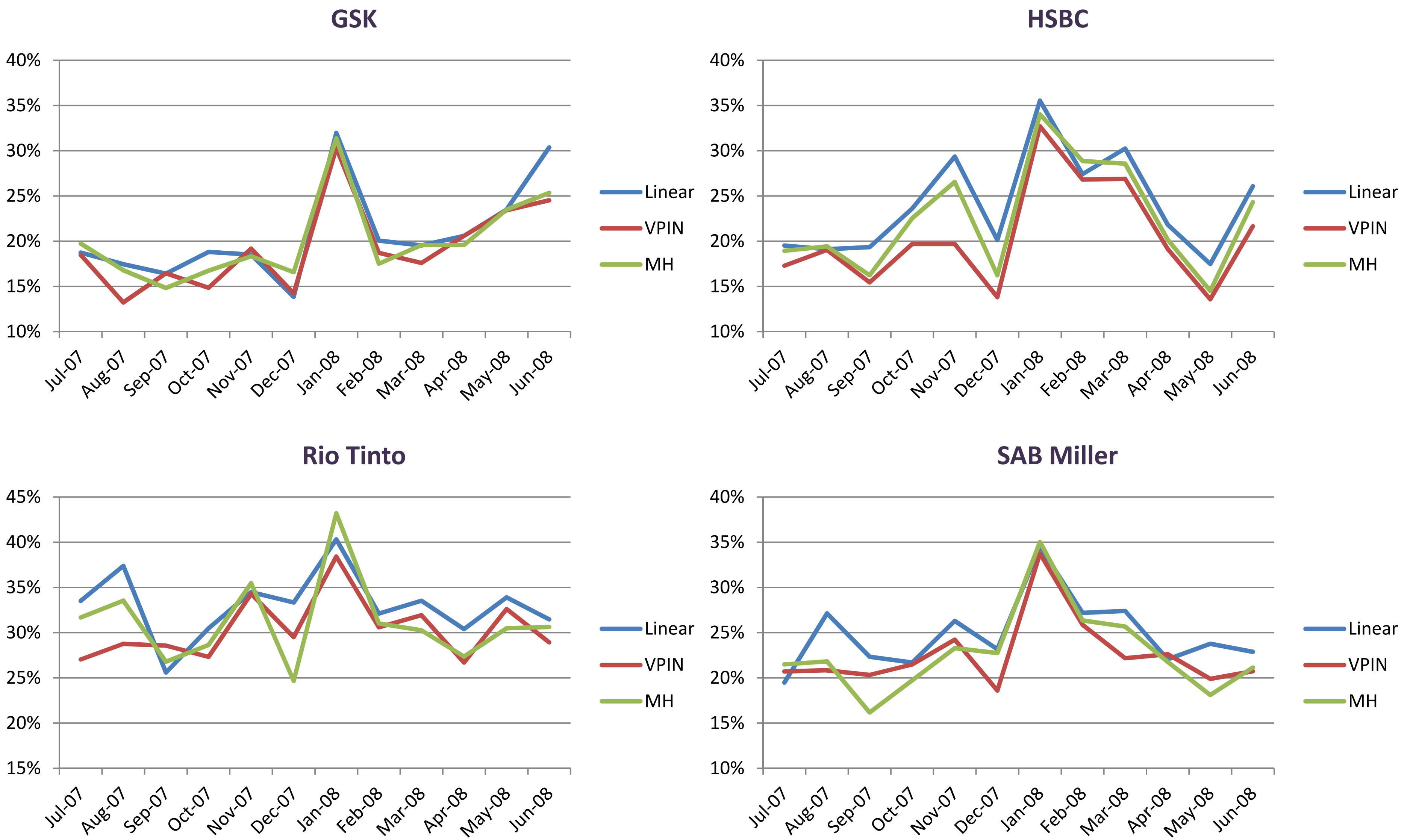
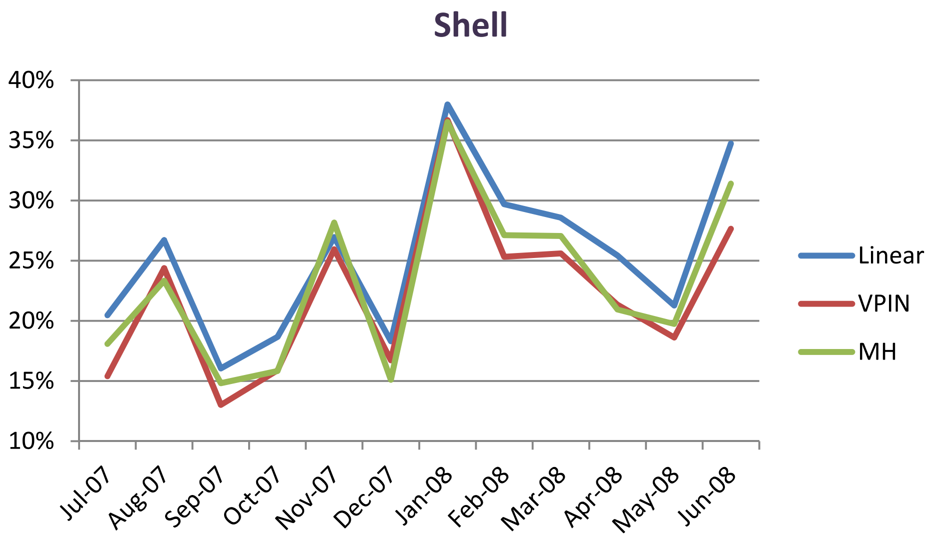


Figure F.1.3: Precision for LSE Stocks (Crash Threshold: -0.10%)



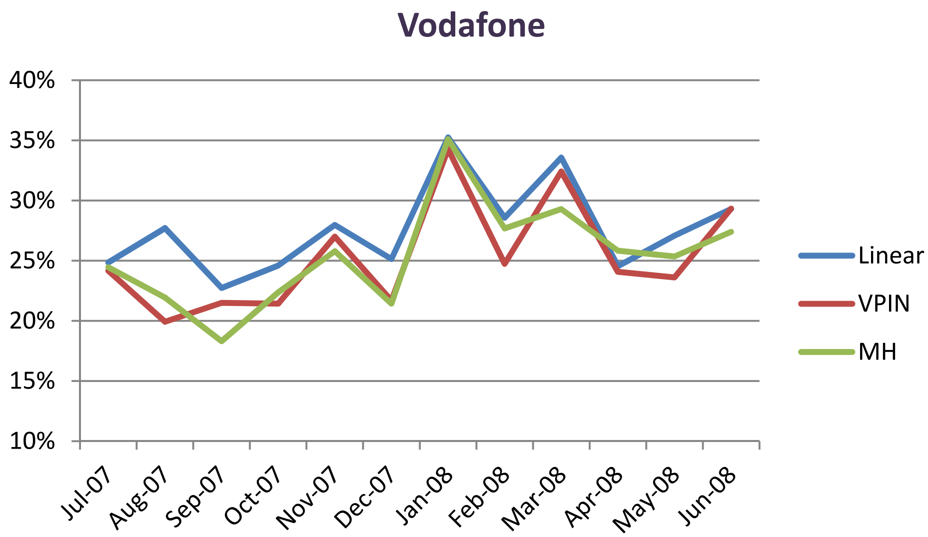


Figure F.2.1: Precision for LSE Stocks (Crash Threshold: -0.25%)

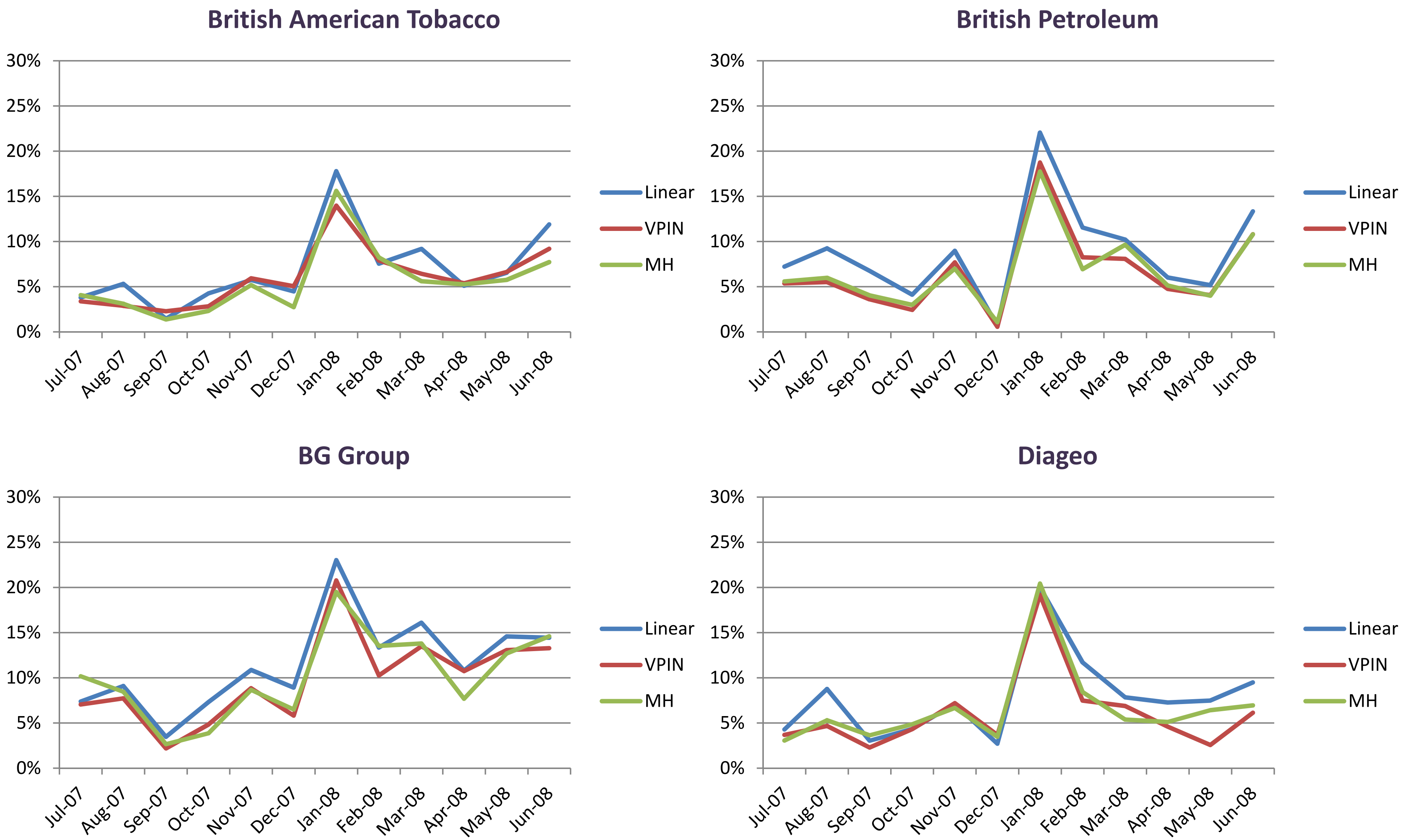


Figure F.2.2: Precision for LSE Stocks (Crash Threshold: -0.25%)

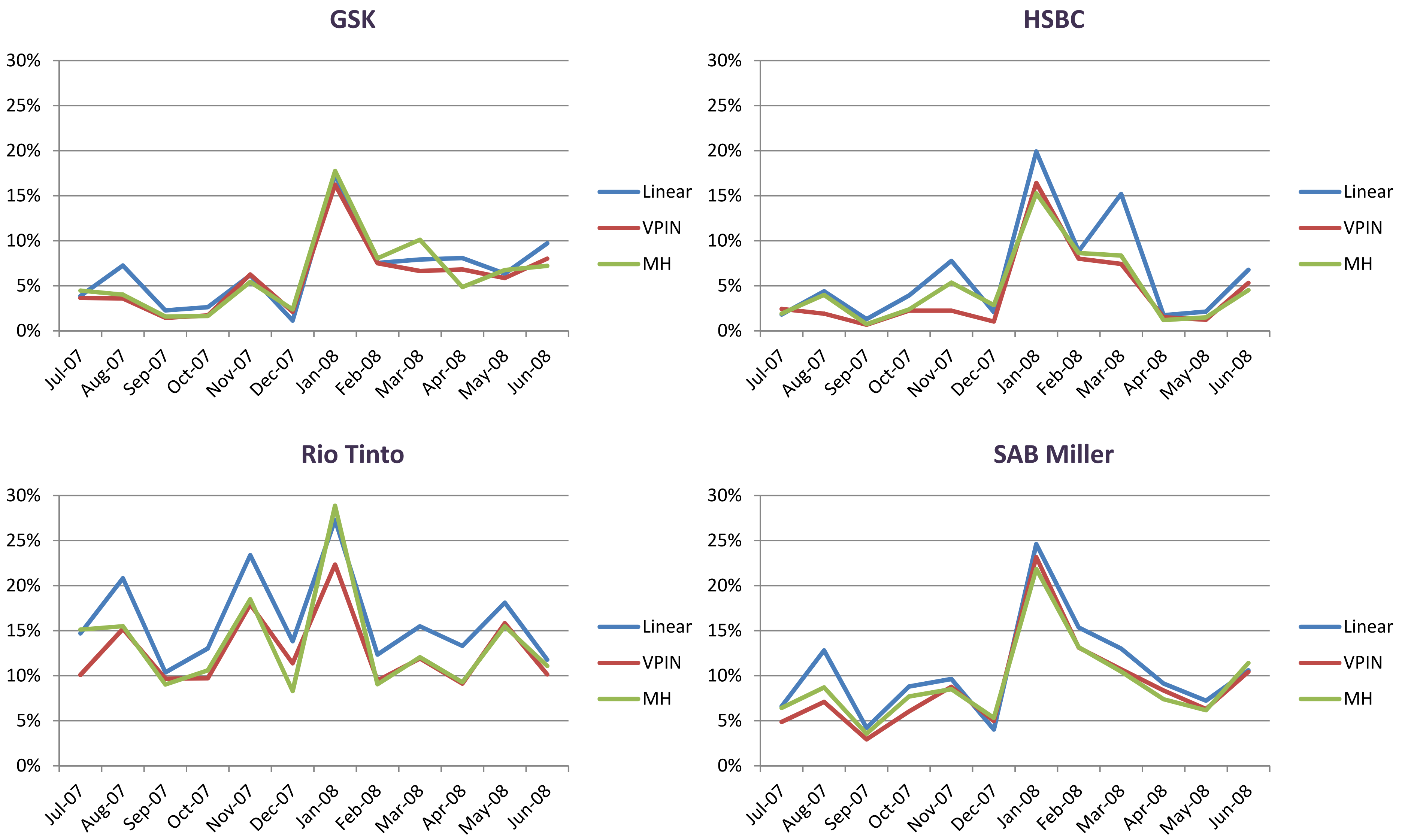


Figure F.2.3: Precision for LSE Stocks (Crash Threshold: -0.25%)

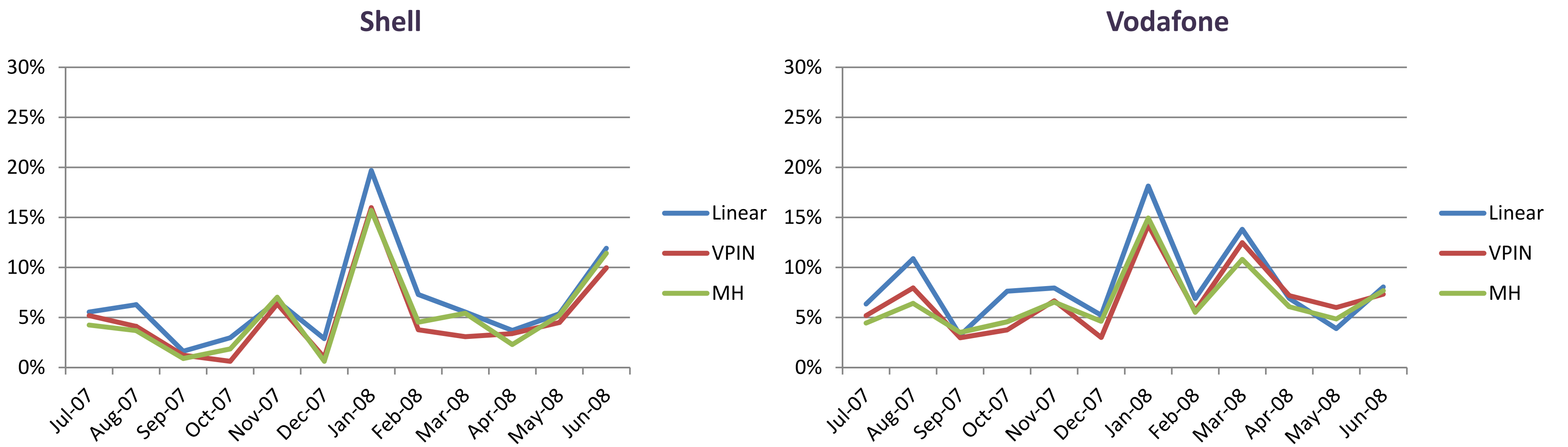


Figure F.3.1: Precision for LSE Stocks (Crash Threshold: -0.50%)

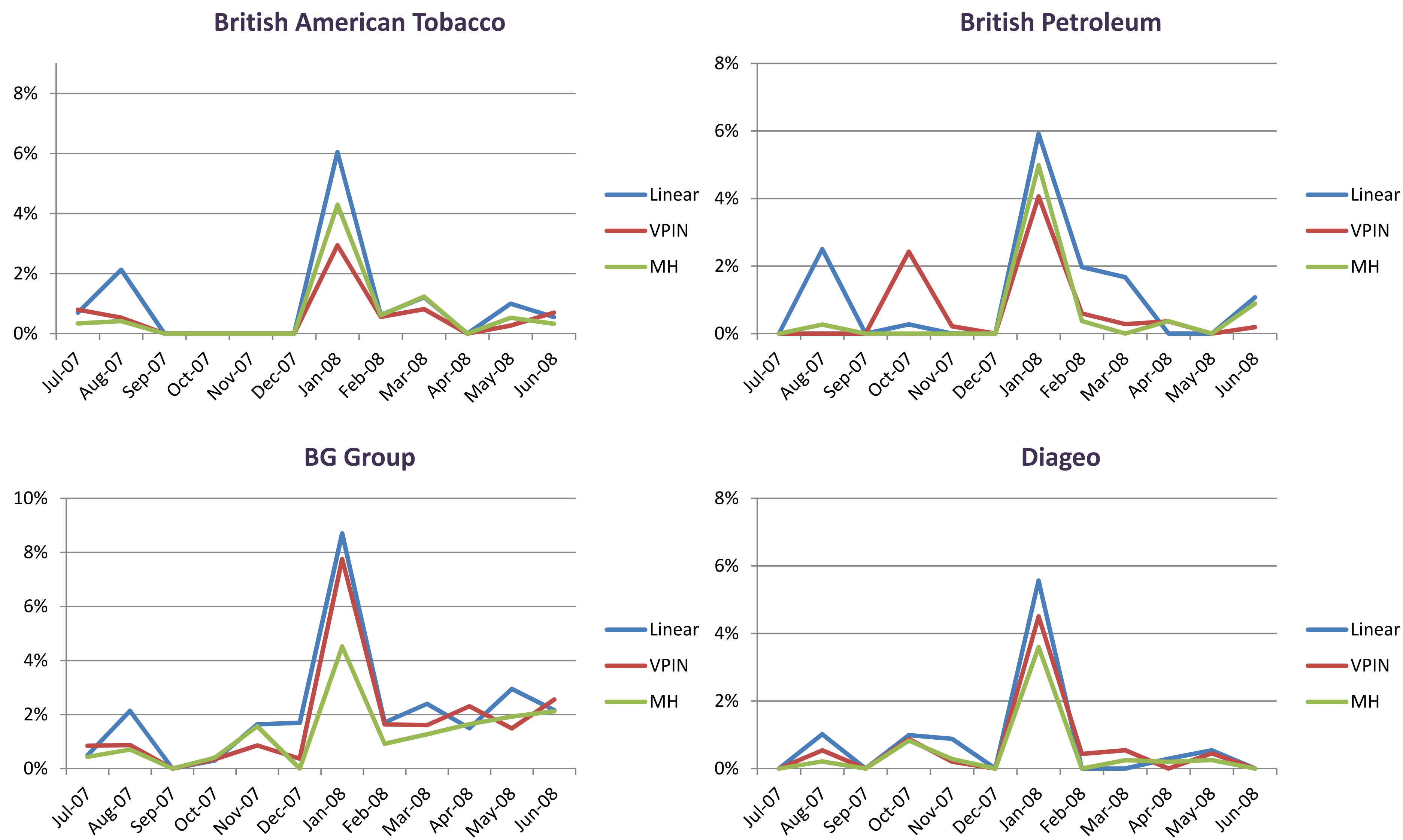


Figure F.3.2: Precision for LSE Stocks (Crash Threshold: -0.50%)

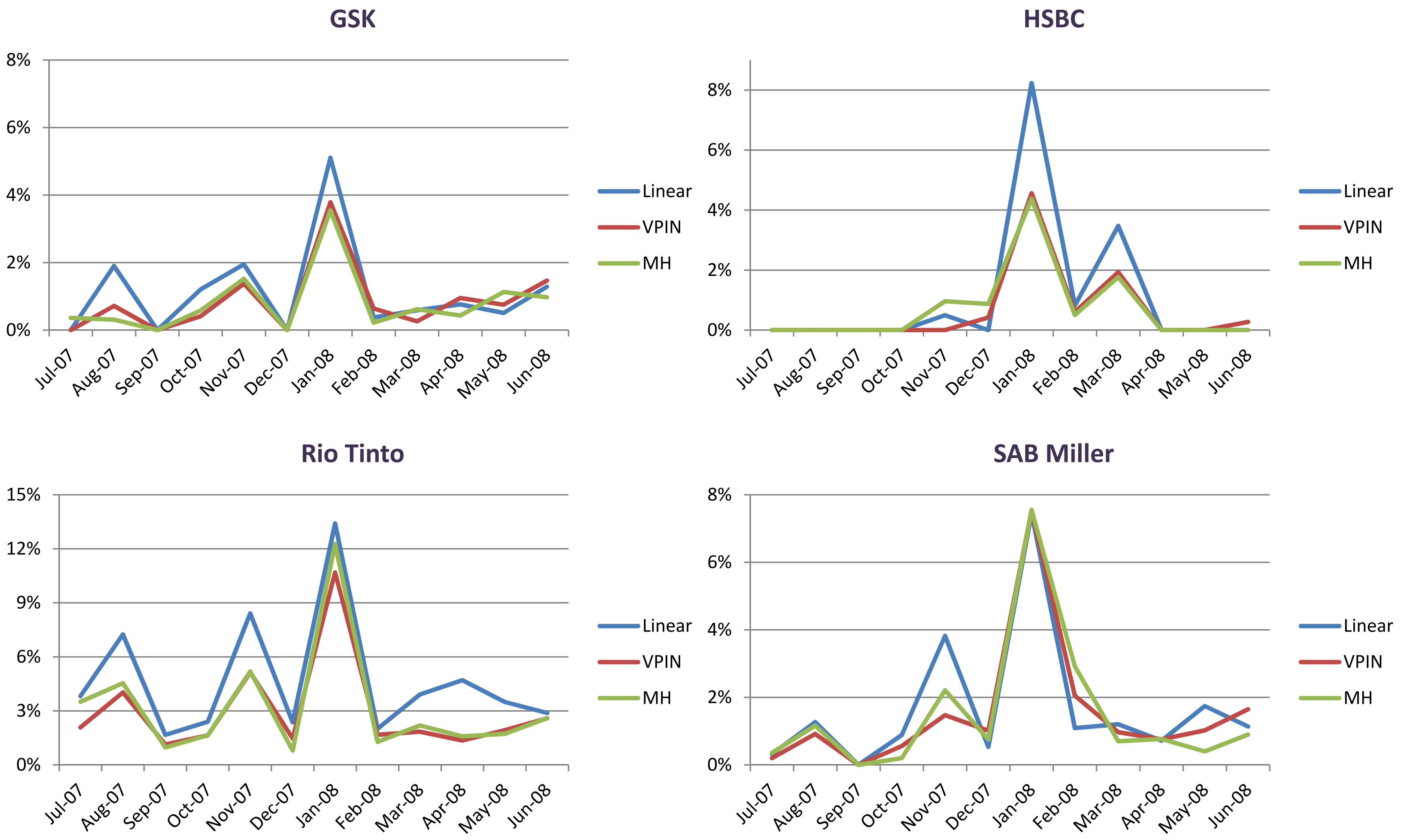
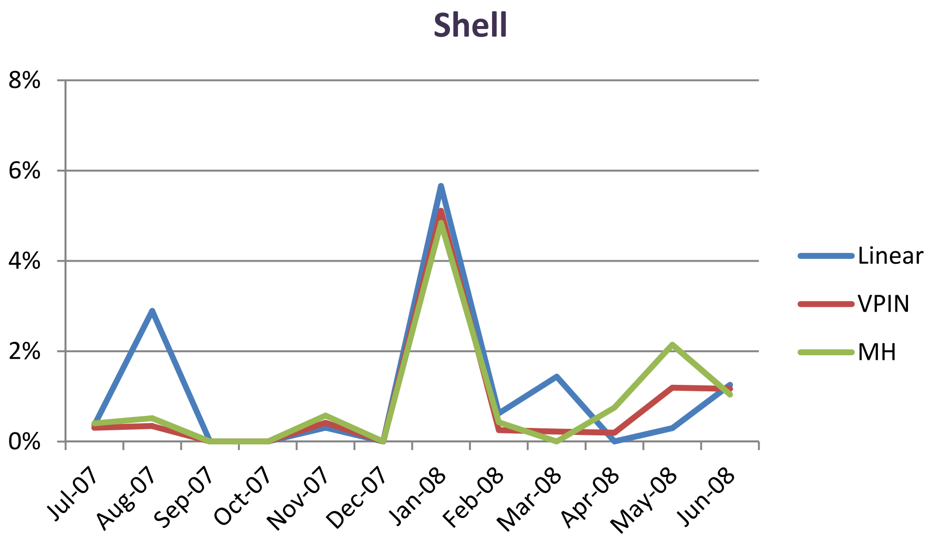


Figure F.3.3: Precision for LSE Stocks (Crash Threshold: -0.50%)



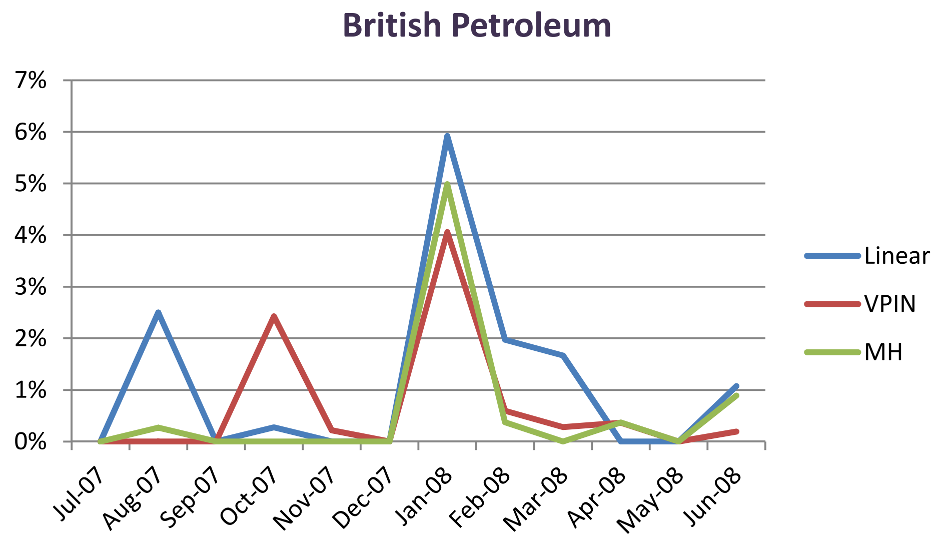


Figure G.1.1: Recall for LSE Stocks (Crash Threshold: -0.10%)

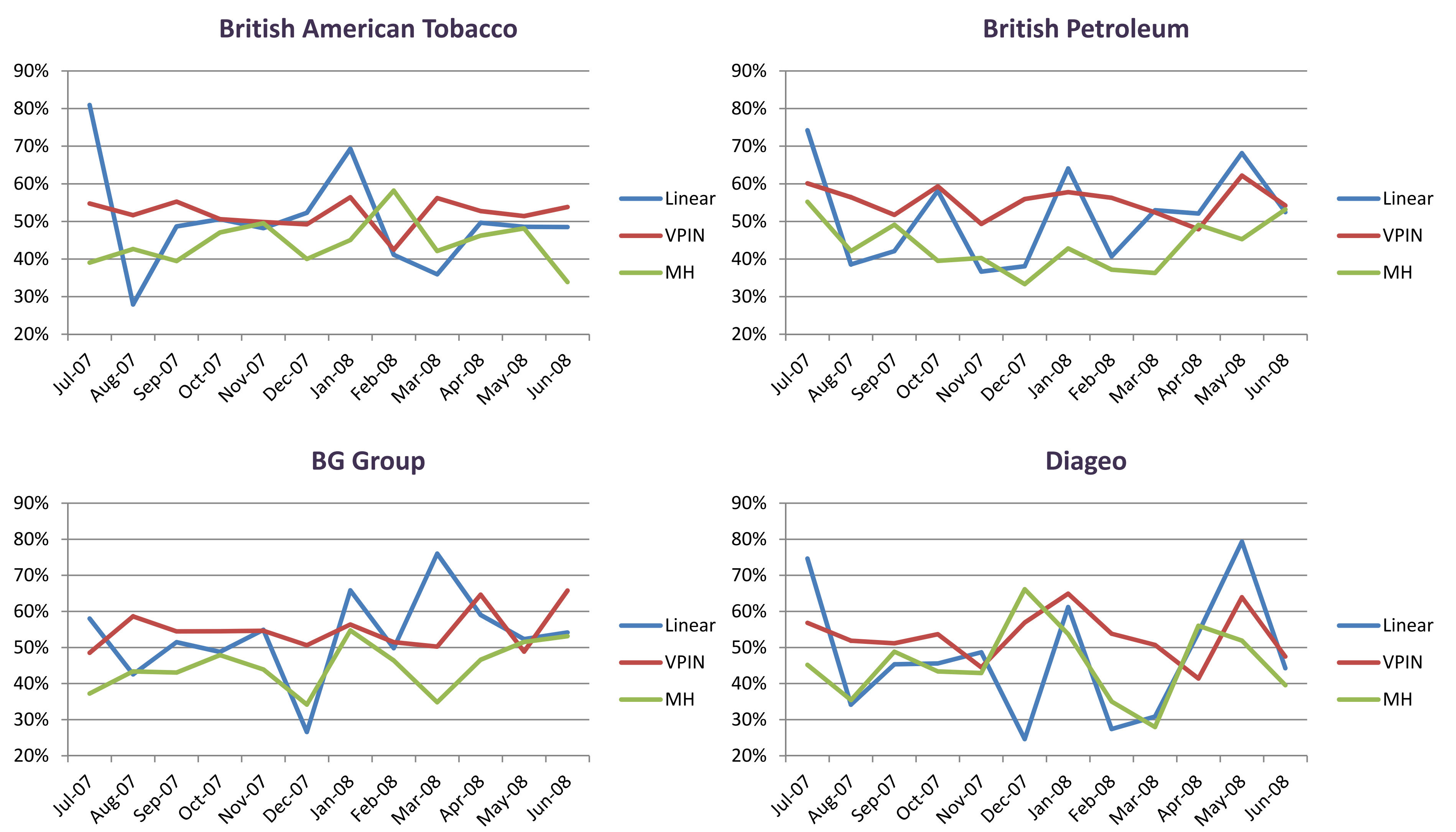


Figure G.1.2: Recall for LSE Stocks (Crash Threshold: -0.10%)

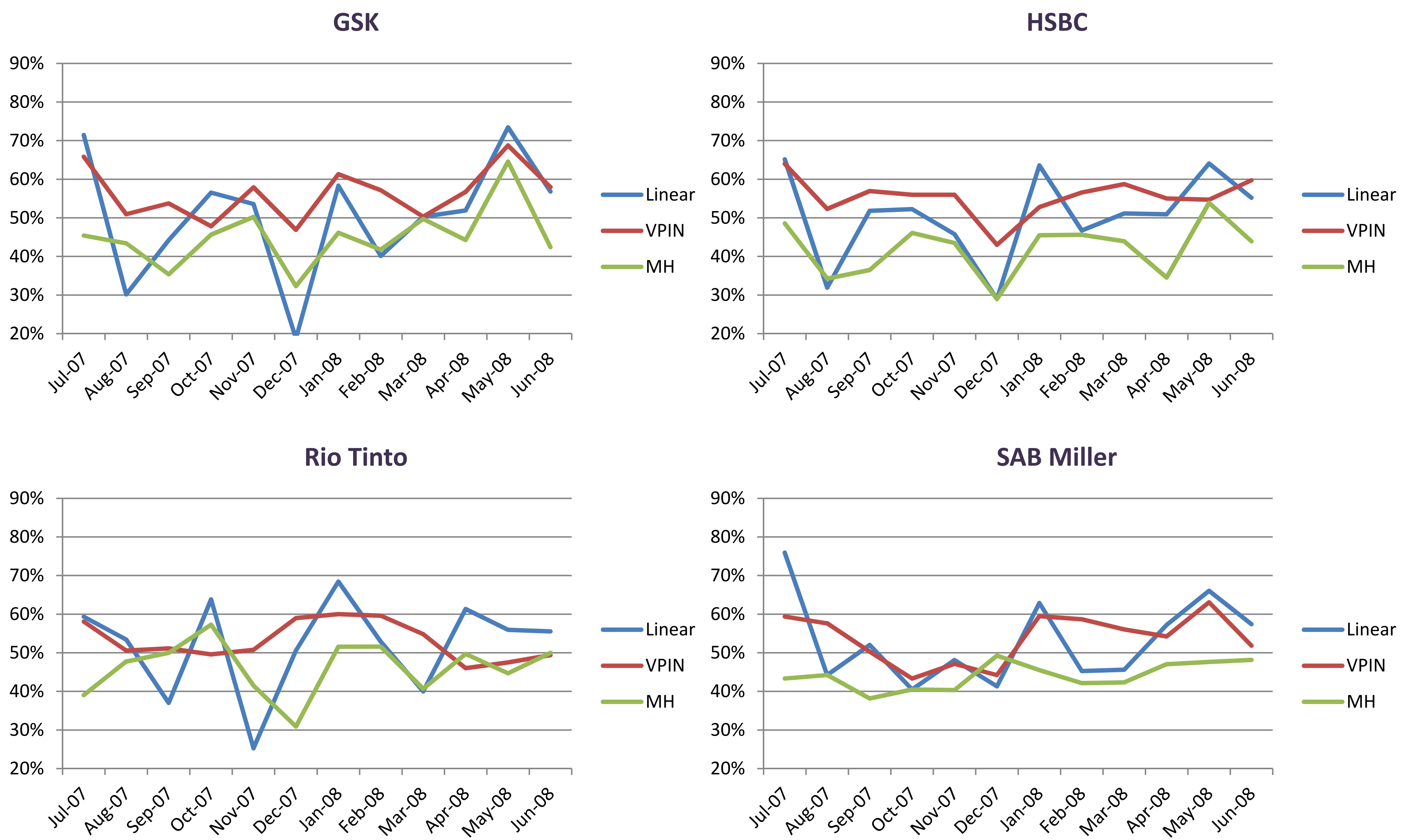
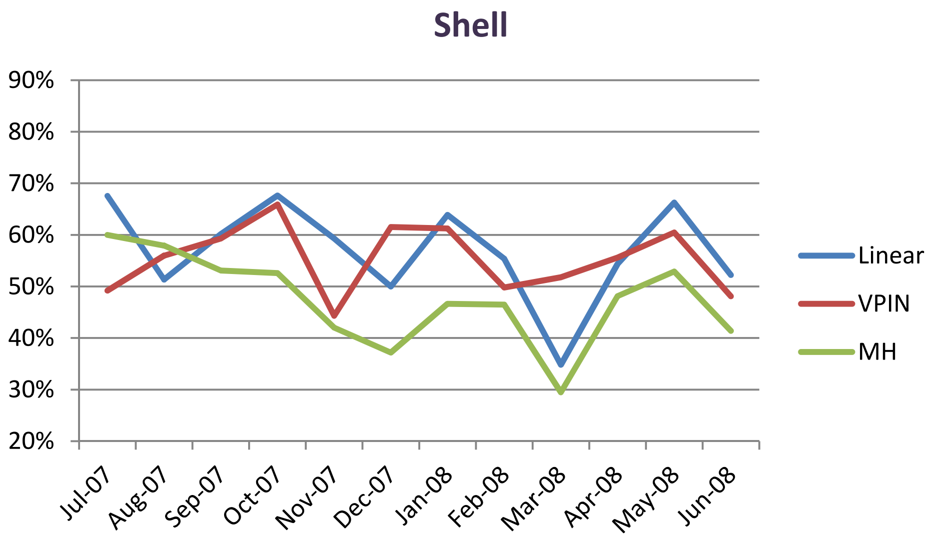


Figure G.1.3: Recall for LSE Stocks (Crash Threshold: -0.10%)



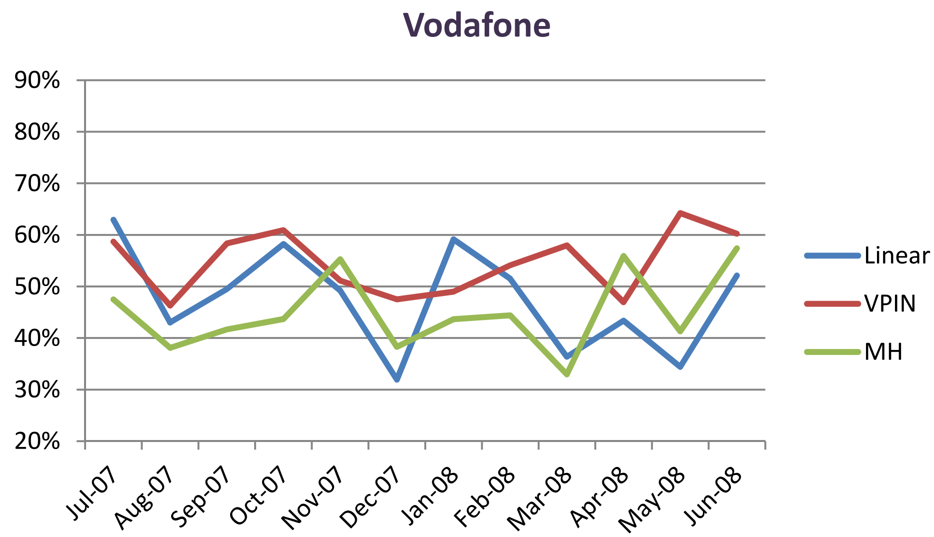


Figure G.2.1: Recall for LSE Stocks (Crash Threshold: -0.25%)

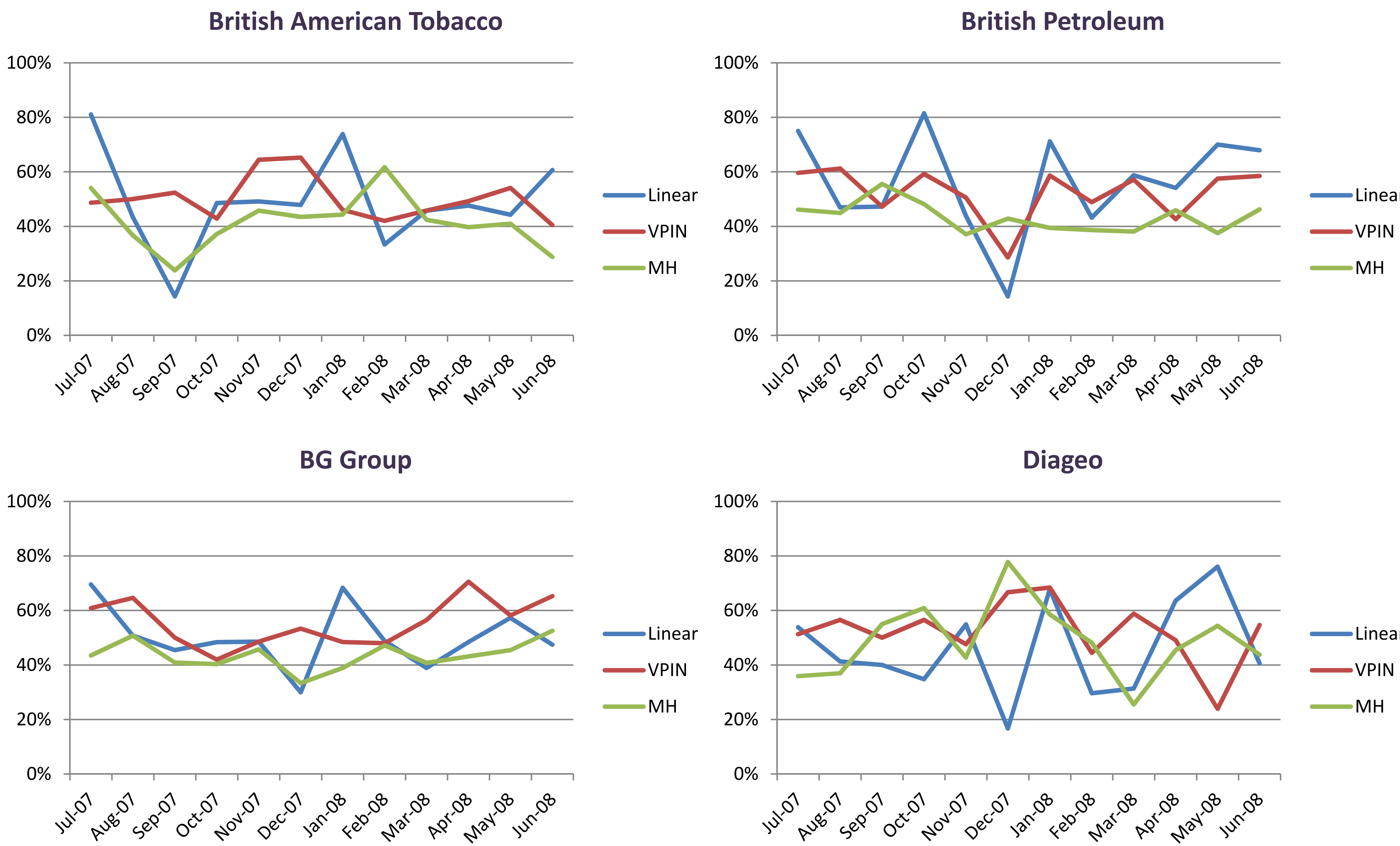


Figure G.2.2: Recall for LSE Stocks (Crash Threshold: -0.25%)

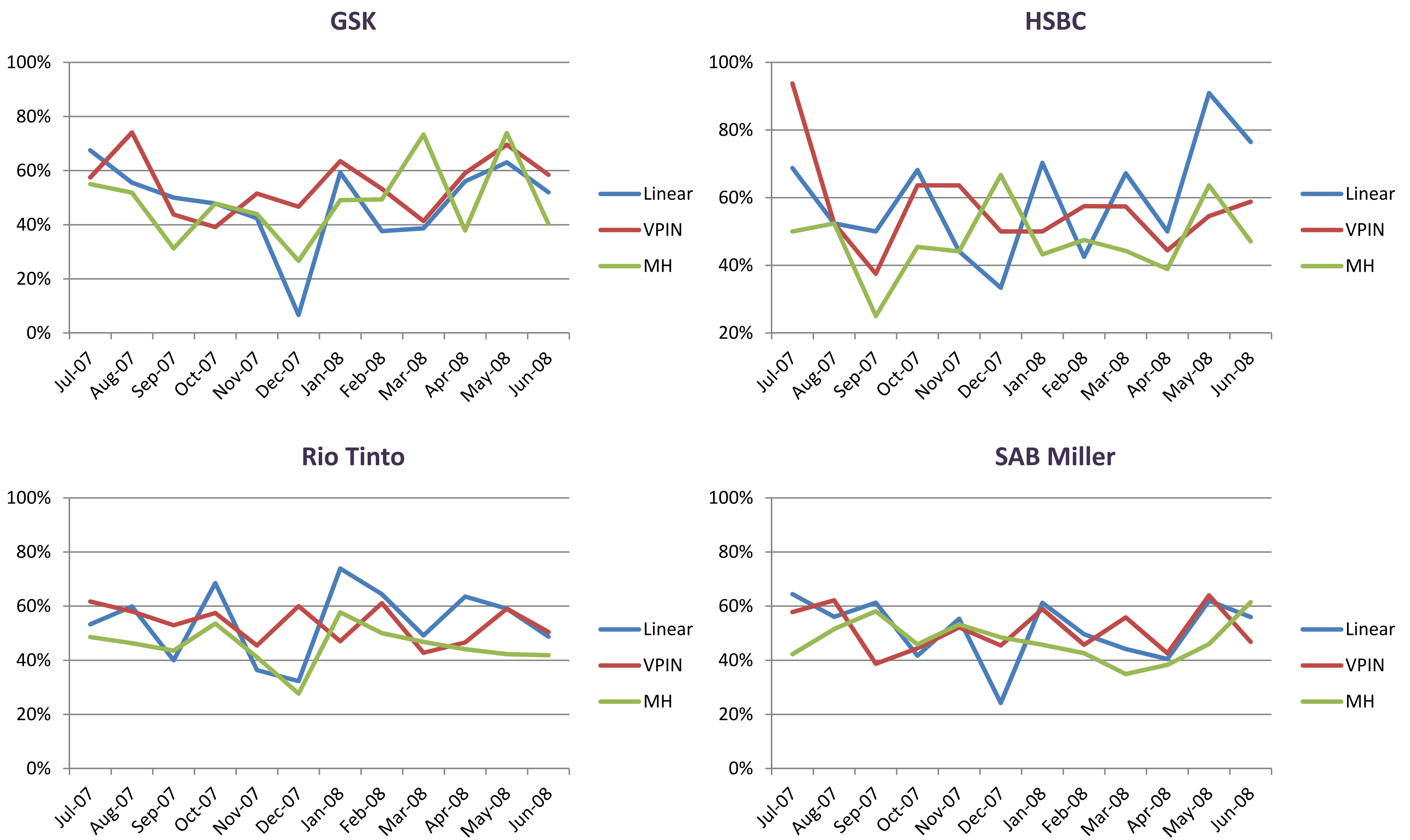


Figure G.2.3: Recall for LSE Stocks (Crash Threshold: -0.25%)

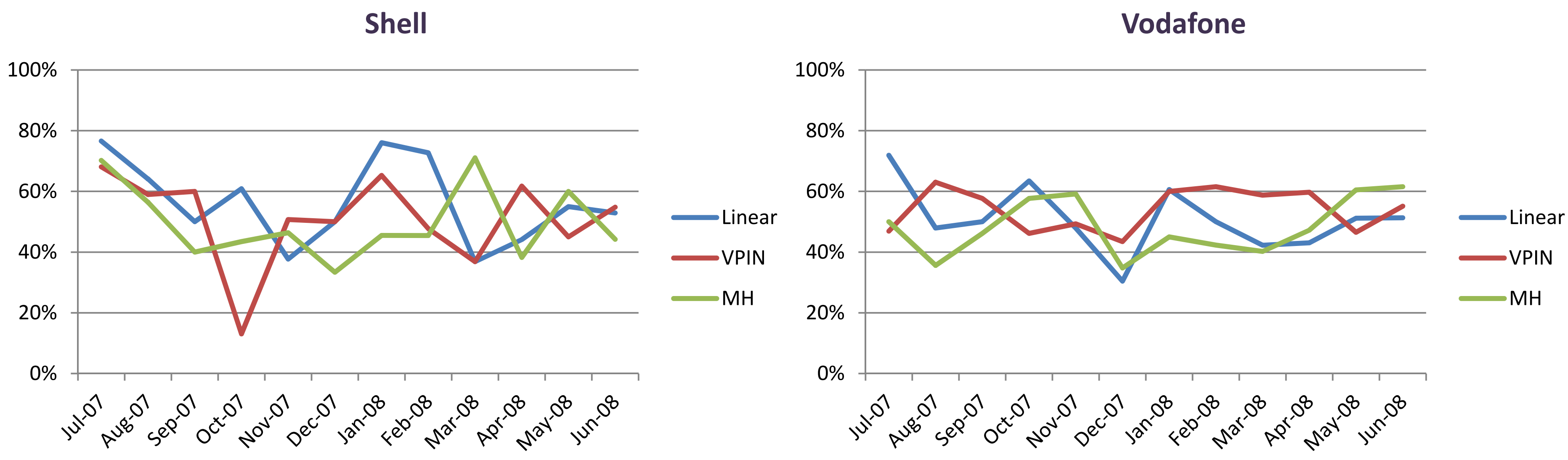


Figure G.3.1: Recall for LSE Stocks (Crash Threshold: -0.50%)

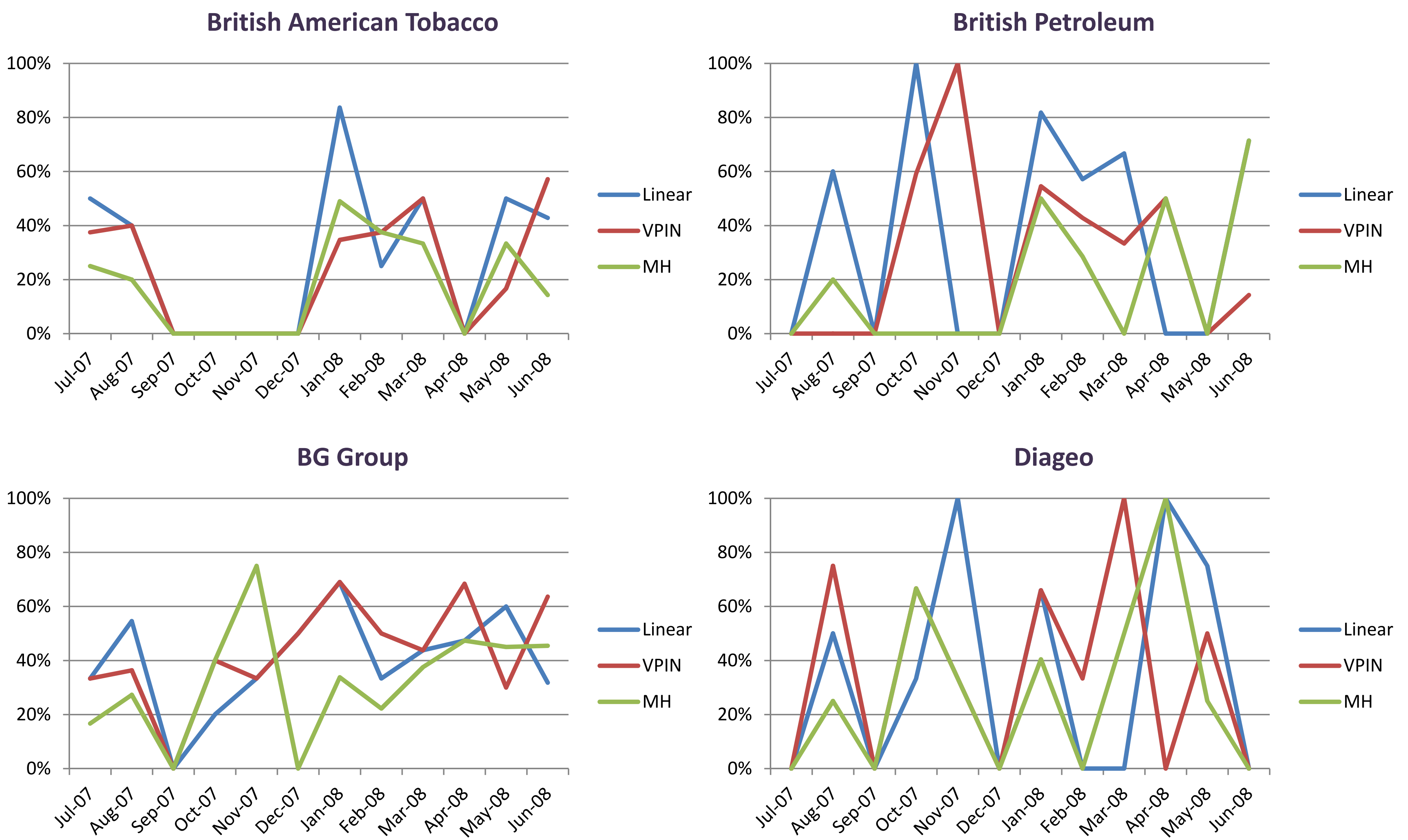


Figure G.3.2: Recall for LSE Stocks (Crash Threshold: -0.50%)

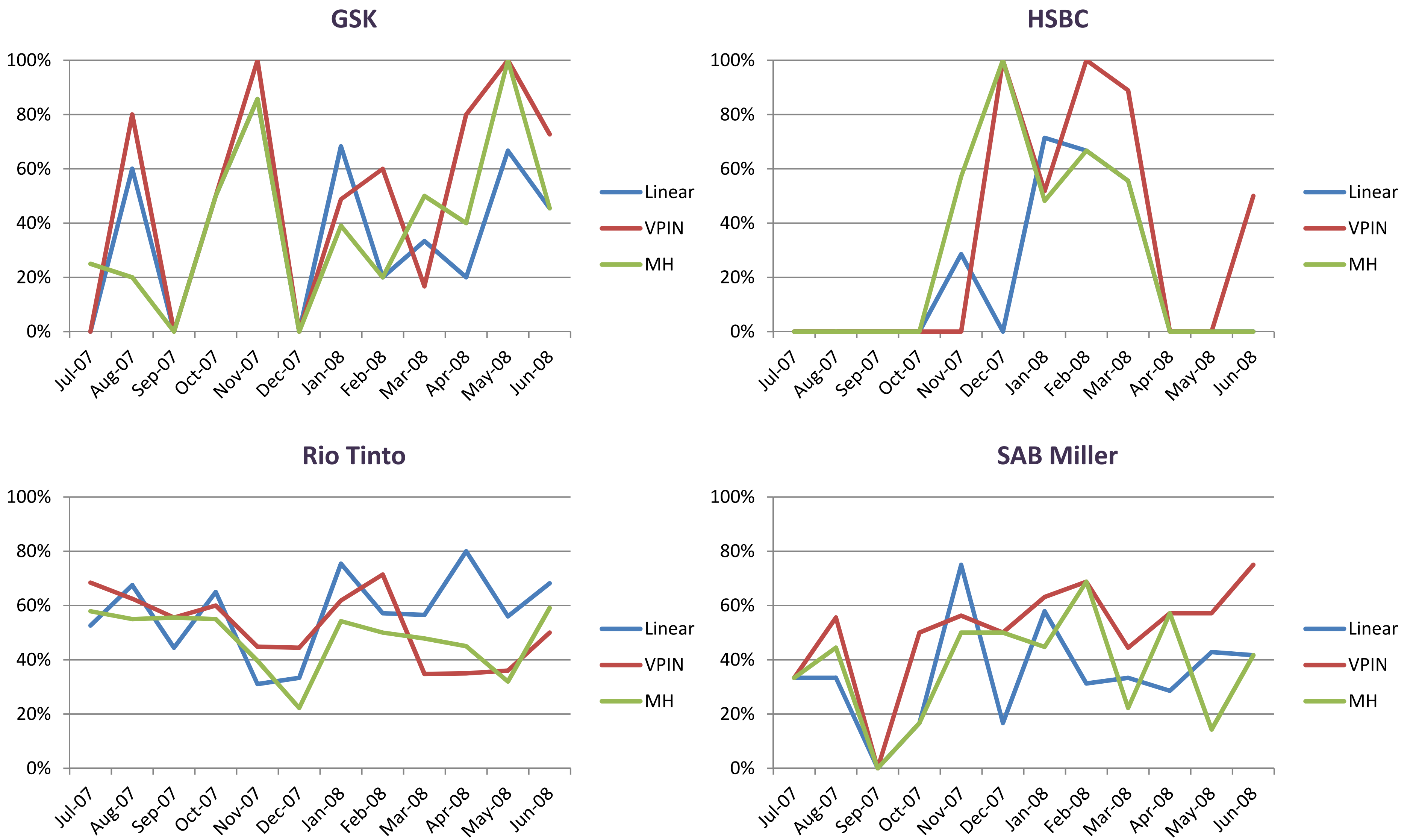
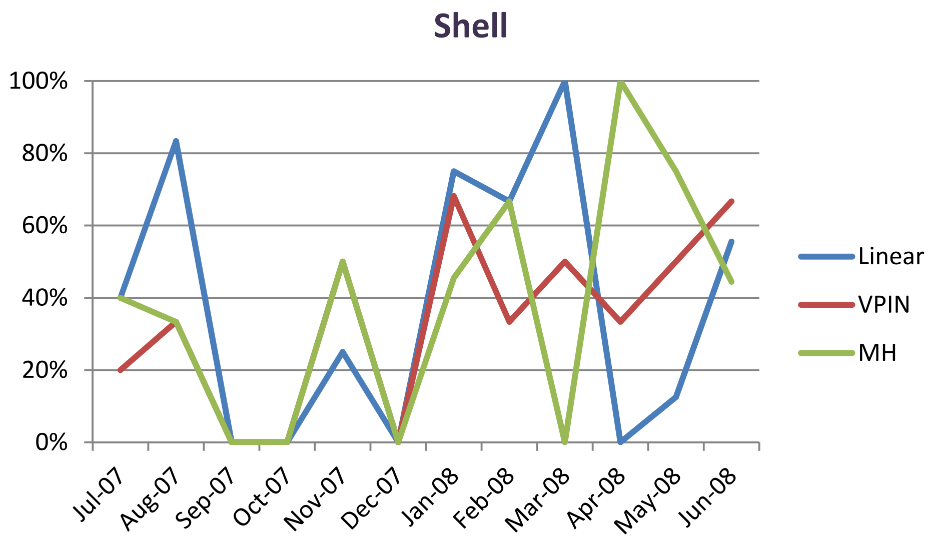
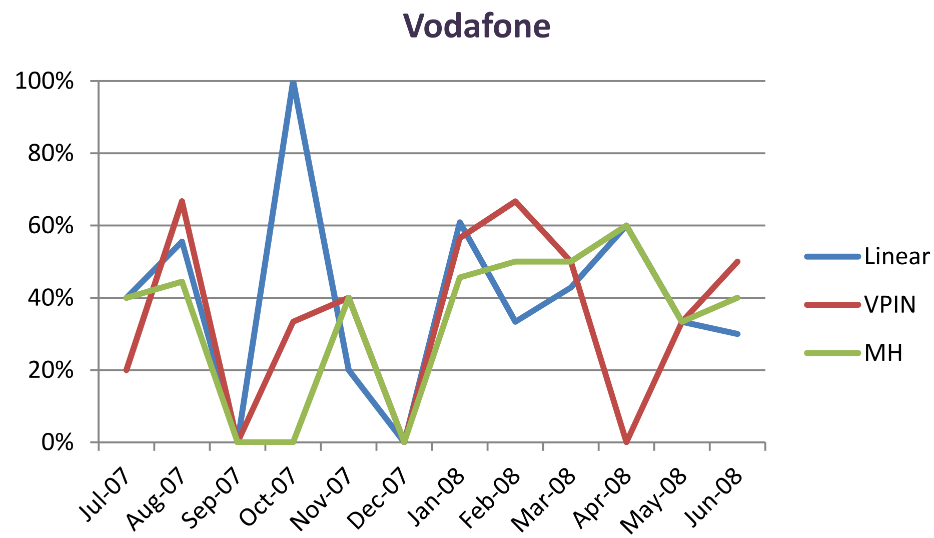


Figure G.3.3: Recall for LSE Stocks (Crash Threshold: -0.50%)





Appendix H

Table H.1.1.1: Out-of-Sample Linear Crash Estimator LDA Confusion Matrices for British American Tobacco Stock (July 2007)

Crash Threshold: -0.10%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	170	40	210
No Crash	693	171	864
Total	863	211	1074

Crash Threshold: -0.25%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	30	7	37
No Crash	759	278	1037
Total	789	285	1074

Crash Threshold: -0.50%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	4	4	8
No Crash	568	498	1066
Total	572	502	1074

Table H.1.1.2: Out-of-Sample Linear Crash Estimator LDA Confusion Matrices for British American Tobacco Stock (August 2007)

Crash Threshold: -0.10%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	59	152	211
No Crash	201	691	892
Total	260	843	1103

Crash Threshold: -0.25%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	13	17	30
No Crash	232	841	1073
Total	245	858	1103

Crash Threshold: -0.50%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	2	3	5
No Crash	92	1006	1098
Total	94	1009	1103

Table H.1.1.3: Out-of-Sample Linear Crash Estimator LDA Confusion Matrices for British American Tobacco Stock (September 2007)

Crash Threshold: -0.10%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	74	78	152
No Crash	298	448	746
Total	372	526	898

Crash Threshold: -0.25%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	3	18	21
No Crash	204	673	877
Total	207	691	898

Crash Threshold: -0.50%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	0	1	1
No Crash	148	749	897
Total	148	750	898

Table H.1.1.4: Out-of-Sample Linear Crash Estimator LDA Confusion Matrices for British American Tobacco Stock (October 2007)

Crash Threshold: -0.10%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	130	127	257
No Crash	443	510	953
Total	573	637	1210

Crash Threshold: -0.25%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	17	18	35
No Crash	382	793	1175
Total	399	811	1210

Crash Threshold: -0.50%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	0	1	1
No Crash	35	1174	1209
Total	35	1175	1210

Table H.1.1.5: Out-of-Sample Linear Crash Estimator LDA Confusion Matrices for British American Tobacco Stock (November 2007)

Crash Threshold: -0.10%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	148	159	307
No Crash	404	429	833
Total	552	588	1140

Crash Threshold: -0.25%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	29	30	59
No Crash	477	604	1081
Total	506	634	1140

Crash Threshold: -0.50%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	0	0	0
No Crash	60	1080	1140
Total	60	1080	1140

Table H.1.1.6: Out-of-Sample Linear Crash Estimator LDA Confusion Matrices for British American Tobacco Stock (December 2007)

Crash Threshold: -0.10%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	68	62	130
No Crash	175	327	502
Total	243	389	632

Crash Threshold: -0.25%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	11	12	23
No Crash	235	374	609
Total	246	386	632

Crash Threshold: -0.50%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	0	1	1
No Crash	197	434	631
Total	197	435	632

Table H.1.1.7: Out-of-Sample Linear Crash Estimator LDA Confusion Matrices for British American Tobacco Stock (January 2008)

Crash Threshold: -0.10%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	280	124	404
No Crash	482	268	750
Total	762	392	1154

Crash Threshold: -0.25%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	130	46	176
No Crash	601	377	978
Total	731	423	1154

Crash Threshold: -0.50%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	41	8	49
No Crash	637	468	1105
Total	678	476	1154

Table H.1.1.8: Out-of-Sample Linear Crash Estimator LDA Confusion Matrices for British American Tobacco Stock (February 2008)

Crash Threshold: -0.10%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	123	176	299
No Crash	301	456	757
Total	424	632	1056

Crash Threshold: -0.25%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	27	54	81
No Crash	331	644	975
Total	358	698	1056

Crash Threshold: -0.50%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	2	6	8
No Crash	319	729	1048
Total	321	735	1056

Table H.1.1.9: Out-of-Sample Linear Crash Estimator LDA Confusion Matrices for British American Tobacco Stock (March 2008)

Crash Threshold: -0.10%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	87	155	242
No Crash	204	428	632
Total	291	583	874

Crash Threshold: -0.25%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	27	32	59
No Crash	267	548	815
Total	294	580	874

Crash Threshold: -0.50%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	3	3	6
No Crash	245	623	868
Total	248	626	874

Table H.1.1.10: Out-of-Sample Linear Crash Estimator LDA Confusion Matrices for British American Tobacco Stock (April 2008)

Crash Threshold: -0.10%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	144	146	290
No Crash	433	418	851
Total	577	564	1141

Crash Threshold: -0.25%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	30	33	63
No Crash	559	519	1078
Total	589	552	1141

Crash Threshold: -0.50%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	0	1	1
No Crash	120	1020	1140
Total	120	1021	1141

Table H.1.1.11: Out-of-Sample Linear Crash Estimator LDA Confusion Matrices for British American Tobacco Stock (May 2008)

Crash Threshold: -0.10%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	119	126	245
No Crash	271	427	698
Total	390	553	943

Crash Threshold: -0.25%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	27	34	61
No Crash	386	496	882
Total	413	530	943

Crash Threshold: -0.50%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	3	3	6
No Crash	297	640	937
Total	300	643	943

Table H.1.1.12: Out-of-Sample Linear Crash Estimator LDA Confusion Matrices for British American Tobacco Stock (June 2008)

Crash Threshold: -0.10%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	146	155	301
No Crash	323	419	742
Total	469	574	1043

Crash Threshold: -0.25%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	57	37	94
No Crash	422	527	949
Total	479	564	1043

Crash Threshold: -0.50%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	3	4	7
No Crash	539	497	1036
Total	542	501	1043

Table H.1.2.1: Out-of-Sample VPIN LDA Confusion Matrices for British American Tobacco Stock (July 2007)

Crash Threshold: -0.10%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	115	95	210
No Crash	468	396	864
Total	583	491	1074

Crash Threshold: -0.25%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	18	19	37
No Crash	515	522	1037
Total	533	541	1074

Crash Threshold: -0.50%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	3	5	8
No Crash	374	692	1066
Total	377	697	1074

Table H.1.2.2: Out-of-Sample VPIN LDA Confusion Matrices for British American Tobacco Stock (August 2007)

Crash Threshold: -0.10%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	109	102	211
No Crash	440	452	892
Total	549	554	1103

Crash Threshold: -0.25%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	15	15	30
No Crash	506	567	1073
Total	521	582	1103

Crash Threshold: -0.50%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	2	3	5
No Crash	373	725	1098
Total	375	728	1103

Table H.1.2.3: Out-of-Sample VPIN LDA Confusion Matrices for British American Tobacco Stock (September 2007)

Crash Threshold: -0.10%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	84	68	152
No Crash	400	346	746
Total	484	414	898

Crash Threshold: -0.25%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	11	10	21
No Crash	471	406	877
Total	482	416	898

Crash Threshold: -0.50%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	0	1	1
No Crash	411	486	897
Total	411	487	898

Table H.1.2.4: Out-of-Sample VPIN LDA Confusion Matrices for British American Tobacco Stock (October 2007)

Crash Threshold: -0.10%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	130	127	257
No Crash	463	490	953
Total	593	617	1210

Crash Threshold: -0.25%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	15	20	35
No Crash	517	658	1175
Total	532	678	1210

Crash Threshold: -0.50%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	0	1	1
No Crash	292	917	1209
Total	292	918	1210

Table H.1.2.5: Out-of-Sample VPIN LDA Confusion Matrices for British American Tobacco Stock (November 2007)

Crash Threshold: -0.10%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	153	154	307
No Crash	466	367	833
Total	619	521	1140

Crash Threshold: -0.25%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	38	21	59
No Crash	604	477	1081
Total	642	498	1140

Crash Threshold: -0.50%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	0	0	0
No Crash	125	1015	1140
Total	125	1015	1140

Table H.1.2.6: Out-of-Sample VPIN LDA Confusion Matrices for British American Tobacco Stock (December 2007)

Crash Threshold: -0.10%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	64	66	130
No Crash	252	250	502
Total	316	316	632

Crash Threshold: -0.25%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	15	8	23
No Crash	282	327	609
Total	297	335	632

Crash Threshold: -0.50%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	0	1	1
No Crash	307	324	631
Total	307	325	632

Table H.1.2.7: Out-of-Sample VPIN LDA Confusion Matrices for British American Tobacco Stock (January 2008)

Crash Threshold: -0.10%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	228	176	404
No Crash	417	333	750
Total	645	509	1154

Crash Threshold: -0.25%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	81	95	176
No Crash	499	479	978
Total	580	574	1154

Crash Threshold: -0.50%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	17	32	49
No Crash	561	544	1105
Total	578	576	1154

Table H.1.2.8: Out-of-Sample VPIN LDA Confusion Matrices for British American Tobacco Stock (February 2008)

Crash Threshold: -0.10%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	127	172	299
No Crash	328	429	757
Total	455	601	1056

Crash Threshold: -0.25%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	34	47	81
No Crash	397	578	975
Total	431	625	1056

Crash Threshold: -0.50%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	3	5	8
No Crash	532	516	1048
Total	535	521	1056

Table H.1.2.9: Out-of-Sample VPIN LDA Confusion Matrices for British American Tobacco Stock (March 2008)

Crash Threshold: -0.10%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	136	106	242
No Crash	341	291	632
Total	477	397	874

Crash Threshold: -0.25%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	27	32	59
No Crash	393	422	815
Total	420	454	874

Crash Threshold: -0.50%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	3	3	6
No Crash	365	503	868
Total	368	506	874

Table H.1.2.10: Out-of-Sample VPIN LDA Confusion Matrices for British American Tobacco Stock (April 2008)

Crash Threshold: -0.10%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	153	137	290
No Crash	445	406	851
Total	598	543	1141

Crash Threshold: -0.25%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	31	32	63
No Crash	546	532	1078
Total	577	564	1141

Crash Threshold: -0.50%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	0	1	1
No Crash	350	790	1140
Total	350	791	1141

Table H.1.2.11: Out-of-Sample VPIN LDA Confusion Matrices for British American Tobacco Stock (May 2008)

Crash Threshold: -0.10%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	126	119	245
No Crash	353	345	698
Total	479	464	943

Crash Threshold: -0.25%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	33	28	61
No Crash	465	417	882
Total	498	445	943

Crash Threshold: -0.50%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	1	5	6
No Crash	369	568	937
Total	370	573	943

Table H.1.2.12: Out-of-Sample VPIN LDA Confusion Matrices for British American Tobacco Stock (June 2008)

Crash Threshold: -0.10%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	162	139	301
No Crash	414	328	742
Total	576	467	1043

Crash Threshold: -0.25%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	38	56	94
No Crash	375	574	949
Total	413	630	1043

Crash Threshold: -0.50%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	4	3	7
No Crash	566	470	1036
Total	570	473	1043

Table H.1.3.1: Out-of-Sample Market Heat LDA Confusion Matrices for British American Tobacco Stock (July 2007)

Crash Threshold: -0.10%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	82	128	210
No Crash	398	466	864
Total	480	594	1074

Crash Threshold: -0.25%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	20	17	37
No Crash	471	566	1037
Total	491	583	1074

Crash Threshold: -0.50%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	2	6	8
No Crash	586	480	1066
Total	588	486	1074

Table H.1.3.2: Out-of-Sample Market Heat LDA Confusion Matrices for British American Tobacco Stock (August 2007)

Crash Threshold: -0.10%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	90	121	211
No Crash	275	617	892
Total	365	738	1103

Crash Threshold: -0.25%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	11	19	30
No Crash	343	730	1073
Total	354	749	1103

Crash Threshold: -0.50%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	1	4	5
No Crash	239	859	1098
Total	240	863	1103

Table H.1.3.3: Out-of-Sample Market Heat LDA Confusion Matrices for British American Tobacco Stock (September 2007)

Crash Threshold: -0.10%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	60	92	152
No Crash	215	531	746
Total	275	623	898

Crash Threshold: -0.25%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	5	16	21
No Crash	358	519	877
Total	363	535	898

Crash Threshold: -0.50%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	0	1	1
No Crash	374	523	897
Total	374	524	898

Table H.1.3.4: Out-of-Sample Market Heat LDA Confusion Matrices for British American Tobacco Stock (October 2007)

Crash Threshold: -0.10%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	121	136	257
No Crash	476	477	953
Total	597	613	1210

Crash Threshold: -0.25%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	13	22	35
No Crash	548	627	1175
Total	561	649	1210

Crash Threshold: -0.50%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	0	1	1
No Crash	187	1022	1209
Total	187	1023	1210

Table H.1.3.5: Out-of-Sample Market Heat LDA Confusion Matrices for British American Tobacco Stock (November 2007)

Crash Threshold: -0.10%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	152	155	307
No Crash	425	408	833
Total	577	563	1140

Crash Threshold: -0.25%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	27	32	59
No Crash	495	586	1081
Total	522	618	1140

Crash Threshold: -0.50%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	0	0	0
No Crash	204	936	1140
Total	204	936	1140

Table H.1.3.6: Out-of-Sample Market Heat LDA Confusion Matrices for British American Tobacco Stock (December 2007)

Crash Threshold: -0.10%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	52	78	130
No Crash	134	368	502
Total	186	446	632

Crash Threshold: -0.25%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	10	13	23
No Crash	356	253	609
Total	366	266	632

Crash Threshold: -0.50%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	0	1	1
No Crash	227	404	631
Total	227	405	632

Table H.1.3.7: Out-of-Sample Market Heat LDA Confusion Matrices for British American Tobacco Stock (January 2008)

Crash Threshold: -0.10%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	182	222	404
No Crash	321	429	750
Total	503	651	1154

Crash Threshold: -0.25%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	78	98	176
No Crash	422	556	978
Total	500	654	1154

Crash Threshold: -0.50%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	24	25	49
No Crash	535	570	1105
Total	559	595	1154

Table H.1.3.8: Out-of-Sample Market Heat LDA Confusion Matrices for British American Tobacco Stock (February 2008)

Crash Threshold: -0.10%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	174	125	299
No Crash	433	324	757
Total	607	449	1056

Crash Threshold: -0.25%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	50	31	81
No Crash	557	418	975
Total	607	449	1056

Crash Threshold: -0.50%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	3	5	8
No Crash	489	559	1048
Total	492	564	1056

Table H.1.3.9: Out-of-Sample Market Heat LDA Confusion Matrices for British American Tobacco Stock (March 2008)

Crash Threshold: -0.10%

Actual		Predicted: Crash	Predicted: No Crash	Total
	Crash	102	140	242
	No Crash	268	364	632
	Total	370	504	874

Crash Threshold: -0.25%

Actual		Predicted: Crash	Predicted: No Crash	Total
	Crash	25	34	59
	No Crash	421	394	815
	Total	446	428	874

Crash Threshold: -0.50%

Actual		Predicted: Crash	Predicted: No Crash	Total
	Crash	2	4	6
	No Crash	160	708	868
	Total	162	712	874

Table H.1.3.10: Out-of-Sample Market Heat LDA Confusion Matrices for British American Tobacco Stock (April 2008)

Crash Threshold: -0.10%

Actual		Predicted: Crash	Predicted: No Crash	Total
	Crash	134	156	290
	No Crash	374	477	851
	Total	508	633	1141

Crash Threshold: -0.25%

Actual		Predicted: Crash	Predicted: No Crash	Total
	Crash	25	38	63
	No Crash	450	628	1078
	Total	475	666	1141

Crash Threshold: -0.50%

Actual		Predicted: Crash	Predicted: No Crash	Total
	Crash	0	1	1
	No Crash	394	746	1140
	Total	394	747	1141

Table H.1.3.11: Out-of-Sample Market Heat LDA Confusion Matrices for British American Tobacco Stock (May 2008)

Crash Threshold: -0.10%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	118	127	245
No Crash	329	369	698
Total	447	496	943

Crash Threshold: -0.25%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	25	36	61
No Crash	410	472	882
Total	435	508	943

Crash Threshold: -0.50%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	2	4	6
No Crash	376	561	937
Total	378	565	943

Table H.1.3.12: Out-of-Sample Market Heat LDA Confusion Matrices for British American Tobacco Stock (June 2008)

Crash Threshold: -0.10%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	102	199	301
No Crash	277	465	742
Total	379	664	1043

Crash Threshold: -0.25%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	27	67	94
No Crash	323	626	949
Total	350	693	1043

Crash Threshold: -0.50%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	1	6	7
No Crash	302	734	1036
Total	303	740	1043

Table H.2.1.1: Out-of-Sample Linear Crash Estimator LDA Confusion Matrices for BG Group Stock (July 2007)

Crash Threshold: -0.10%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	159	115	274
No Crash	445	388	833
Total	604	503	1107

Crash Threshold: -0.25%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	48	21	69
No Crash	603	435	1038
Total	651	456	1107

Crash Threshold: -0.50%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	4	8	12
No Crash	800	295	1095
Total	804	303	1107

Table H.2.1.2: Out-of-Sample Linear Crash Estimator LDA Confusion Matrices for BG Group Stock (August 2007)

Crash Threshold: -0.10%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	106	143	249
No Crash	183	703	886
Total	289	846	1135

Crash Threshold: -0.25%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	33	32	65
No Crash	330	740	1070
Total	363	772	1135

Crash Threshold: -0.50%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	6	5	11
No Crash	275	849	1124
Total	281	854	1135

Table H.2.1.3: Out-of-Sample Linear Crash Estimator LDA Confusion Matrices for BG Group Stock (September 2007)

Crash Threshold: -0.10%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	104	98	202
No Crash	330	414	744
Total	434	512	946

Crash Threshold: -0.25%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	10	12	22
No Crash	279	645	924
Total	289	657	946

Crash Threshold: -0.50%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	0	1	1
No Crash	95	850	945
Total	95	851	946

Table H.2.1.4: Out-of-Sample Linear Crash Estimator LDA Confusion Matrices for BG Group Stock (October 2007)

Crash Threshold: -0.10%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	162	170	332
No Crash	371	533	904
Total	533	703	1236

Crash Threshold: -0.25%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	30	32	62
No Crash	380	794	1174
Total	410	826	1236

Crash Threshold: -0.50%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	1	4	5
No Crash	330	901	1231
Total	331	905	1236

Table H.2.1.5: Out-of-Sample Linear Crash Estimator LDA Confusion Matrices for BG Group Stock (November 2007)

Crash Threshold: -0.10%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	190	156	346
No Crash	390	413	803
Total	580	569	1149

Crash Threshold: -0.25%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	51	54	105
No Crash	418	626	1044
Total	469	680	1149

Crash Threshold: -0.50%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	4	8	12
No Crash	240	897	1137
Total	244	905	1149

Table H.2.1.6: Out-of-Sample Linear Crash Estimator LDA Confusion Matrices for BG Group Stock (December 2007)

Crash Threshold: -0.10%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	21	58	79
No Crash	90	475	565
Total	111	533	644

Crash Threshold: -0.25%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	9	21	30
No Crash	92	522	614
Total	101	543	644

Crash Threshold: -0.50%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	1	1	2
No Crash	58	584	642
Total	59	585	644

Table H.2.1.7: Out-of-Sample Linear Crash Estimator LDA Confusion Matrices for BG Group Stock (January 2008)

Crash Threshold: -0.10%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	243	126	369
No Crash	442	345	787
Total	685	471	1156

Crash Threshold: -0.25%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	151	70	221
No Crash	505	430	935
Total	656	500	1156

Crash Threshold: -0.50%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	49	22	71
No Crash	514	571	1085
Total	563	593	1156

Table H.2.1.8: Out-of-Sample Linear Crash Estimator LDA Confusion Matrices for BG Group Stock (February 2008)

Crash Threshold: -0.10%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	118	119	237
No Crash	317	507	824
Total	435	626	1061

Crash Threshold: -0.25%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	61	64	125
No Crash	396	540	936
Total	457	604	1061

Crash Threshold: -0.50%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	6	12	18
No Crash	345	698	1043
Total	351	710	1061

Table H.2.1.9: Out-of-Sample Linear Crash Estimator LDA Confusion Matrices for BG Group Stock (March 2008)

Crash Threshold: -0.10%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	92	29	121
No Crash	282	468	750
Total	374	497	871

Crash Threshold: -0.25%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	42	66	108
No Crash	219	544	763
Total	261	610	871

Crash Threshold: -0.50%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	7	9	16
No Crash	285	570	855
Total	292	579	871

Table H.2.1.10: Out-of-Sample Linear Crash Estimator LDA Confusion Matrices for BG Group Stock (April 2008)

Crash Threshold: -0.10%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	157	109	266
No Crash	485	382	867
Total	642	491	1133

Crash Threshold: -0.25%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	46	49	95
No Crash	380	658	1038
Total	426	707	1133

Crash Threshold: -0.50%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	9	10	19
No Crash	594	520	1114
Total	603	530	1133

Table H.2.1.11: Out-of-Sample Linear Crash Estimator LDA Confusion Matrices for BG Group Stock (May 2008)

Crash Threshold: -0.10%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	135	123	258
No Crash	298	396	694
Total	433	519	952

Crash Threshold: -0.25%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	63	47	110
No Crash	369	473	842
Total	432	520	952

Crash Threshold: -0.50%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	12	8	20
No Crash	395	537	932
Total	407	545	952

Table H.2.1.12: Out-of-Sample Linear Crash Estimator LDA Confusion Matrices for BG Group Stock (June 2008)

Crash Threshold: -0.10%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	163	138	301
No Crash	340	408	748
Total	503	546	1049

Crash Threshold: -0.25%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	56	62	118
No Crash	332	599	931
Total	388	661	1049

Crash Threshold: -0.50%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	7	15	22
No Crash	317	710	1027
Total	324	725	1049

Table H.2.2.1: Out-of-Sample VPIN LDA Confusion Matrices for BG Group Stock (July 2007)

Crash Threshold: -0.10%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	133	141	274
No Crash	402	431	833
Total	535	572	1107

Crash Threshold: -0.25%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	42	27	69
No Crash	554	484	1038
Total	596	511	1107

Crash Threshold: -0.50%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	4	8	12
No Crash	470	625	1095
Total	474	633	1107

Table H.2.2.2: Out-of-Sample VPIN LDA Confusion Matrices for BG Group Stock (August 2007)

Crash Threshold: -0.10%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	146	103	249
No Crash	423	463	886
Total	569	566	1135

Crash Threshold: -0.25%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	42	23	65
No Crash	502	568	1070
Total	544	591	1135

Crash Threshold: -0.50%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	4	7	11
No Crash	456	668	1124
Total	460	675	1135

Table H.2.2.3: Out-of-Sample VPIN LDA Confusion Matrices for BG Group Stock (September 2007)

Crash Threshold: -0.10%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	110	92	202
No Crash	407	337	744
Total	517	429	946

Crash Threshold: -0.25%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	11	11	22
No Crash	492	432	924
Total	503	443	946

Crash Threshold: -0.50%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	0	1	1
No Crash	429	516	945
Total	429	517	946

Table H.2.2.4: Out-of-Sample VPIN LDA Confusion Matrices for BG Group Stock (October 2007)

Crash Threshold: -0.10%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	181	151	332
No Crash	505	399	904
Total	686	550	1236

Crash Threshold: -0.25%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	26	36	62
No Crash	512	662	1174
Total	538	698	1236

Crash Threshold: -0.50%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	2	3	5
No Crash	570	661	1231
Total	572	664	1236

Table H.2.2.5: Out-of-Sample VPIN LDA Confusion Matrices for BG Group Stock (November 2007)

Crash Threshold: -0.10%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	189	157	346
No Crash	421	382	803
Total	610	539	1149

Crash Threshold: -0.25%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	51	54	105
No Crash	525	519	1044
Total	576	573	1149

Crash Threshold: -0.50%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	4	8	12
No Crash	463	674	1137
Total	467	682	1149

Table H.2.2.6: Out-of-Sample VPIN LDA Confusion Matrices for BG Group Stock (December 2007)

Crash Threshold: -0.10%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	40	39	79
No Crash	241	324	565
Total	281	363	644

Crash Threshold: -0.25%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	16	14	30
No Crash	259	355	614
Total	275	369	644

Crash Threshold: -0.50%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	1	1	2
No Crash	274	368	642
Total	275	369	644

Table H.2.2.7: Out-of-Sample VPIN LDA Confusion Matrices for BG Group Stock (January 2008)

Crash Threshold: -0.10%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	208	161	369
No Crash	438	349	787
Total	646	510	1156

Crash Threshold: -0.25%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	107	114	221
No Crash	408	527	935
Total	515	641	1156

Crash Threshold: -0.50%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	49	22	71
No Crash	583	502	1085
Total	632	524	1156

Table H.2.2.8: Out-of-Sample VPIN LDA Confusion Matrices for BG Group Stock (February 2008)

Crash Threshold: -0.10%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	122	115	237
No Crash	466	358	824
Total	588	473	1061

Crash Threshold: -0.25%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	60	65	125
No Crash	525	411	936
Total	585	476	1061

Crash Threshold: -0.50%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	9	9	18
No Crash	541	502	1043
Total	550	511	1061

Table H.2.2.9: Out-of-Sample VPIN LDA Confusion Matrices for BG Group Stock (March 2008)

Crash Threshold: -0.10%

Actual		Predicted: Crash	Predicted: No Crash	Total
	Crash	101	100	201
	No Crash	352	318	670
	Total	453	418	871

Crash Threshold: -0.25%

Actual		Predicted: Crash	Predicted: No Crash	Total
	Crash	61	47	108
	No Crash	391	372	763
	Total	452	419	871

Crash Threshold: -0.50%

Actual		Predicted: Crash	Predicted: No Crash	Total
	Crash	7	9	16
	No Crash	429	426	855
	Total	436	435	871

Table H.2.2.10: Out-of-Sample VPIN LDA Confusion Matrices for BG Group Stock (April 2008)

Crash Threshold: -0.10%

Actual		Predicted: Crash	Predicted: No Crash	Total
	Crash	172	94	266
	No Crash	470	397	867
	Total	642	491	1133

Crash Threshold: -0.25%

Actual		Predicted: Crash	Predicted: No Crash	Total
	Crash	67	28	95
	No Crash	557	481	1038
	Total	624	509	1133

Crash Threshold: -0.50%

Actual		Predicted: Crash	Predicted: No Crash	Total
	Crash	13	6	19
	No Crash	551	563	1114
	Total	564	569	1133

Table H.2.2.11: Out-of-Sample VPIN LDA Confusion Matrices for BG Group Stock (May 2008)

Crash Threshold: -0.10%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	126	132	258
No Crash	380	314	694
Total	506	446	952

Crash Threshold: -0.25%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	64	46	110
No Crash	426	416	842
Total	490	462	952

Crash Threshold: -0.50%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	6	14	20
No Crash	397	535	932
Total	403	549	952

Table H.2.2.12: Out-of-Sample VPIN LDA Confusion Matrices for BG Group Stock (June 2008)

Crash Threshold: -0.10%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	198	103	301
No Crash	401	347	748
Total	599	450	1049

Crash Threshold: -0.25%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	77	41	118
No Crash	503	428	931
Total	580	469	1049

Crash Threshold: -0.50%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	14	8	22
No Crash	534	493	1027
Total	548	501	1049

Table H.2.3.1: Out-of-Sample Market Heat LDA Confusion Matrices for BG Group Stock (July 2007)

Crash Threshold: -0.10%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	102	172	274
No Crash	187	646	833
Total	289	818	1107

Crash Threshold: -0.25%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	30	39	69
No Crash	265	773	1038
Total	295	812	1107

Crash Threshold: -0.50%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	2	10	12
No Crash	458	637	1095
Total	460	647	1107

Table H.2.3.2: Out-of-Sample Market Heat LDA Confusion Matrices for BG Group Stock (August 2007)

Crash Threshold: -0.10%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	108	141	249
No Crash	303	583	886
Total	411	724	1135

Crash Threshold: -0.25%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	33	32	65
No Crash	357	713	1070
Total	390	745	1135

Crash Threshold: -0.50%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	3	8	11
No Crash	422	702	1124
Total	425	710	1135

Table H.2.3.3: Out-of-Sample Market Heat LDA Confusion Matrices for BG Group Stock (September 2007)

Crash Threshold: -0.10%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	87	115	202
No Crash	281	463	744
Total	368	578	946

Crash Threshold: -0.25%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	9	13	22
No Crash	330	594	924
Total	339	607	946

Crash Threshold: -0.50%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	0	1	1
No Crash	209	736	945
Total	209	737	946

Table H.2.3.4: Out-of-Sample Market Heat LDA Confusion Matrices for BG Group Stock (October 2007)

Crash Threshold: -0.10%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	159	173	332
No Crash	442	462	904
Total	601	635	1236

Crash Threshold: -0.25%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	25	37	62
No Crash	622	552	1174
Total	647	589	1236

Crash Threshold: -0.50%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	2	3	5
No Crash	490	741	1231
Total	492	744	1236

Table H.2.3.5: Out-of-Sample Market Heat LDA Confusion Matrices for BG Group Stock (November 2007)

Crash Threshold: -0.10%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	152	194	346
No Crash	344	459	803
Total	496	653	1149

Crash Threshold: -0.25%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	48	57	105
No Crash	506	538	1044
Total	554	595	1149

Crash Threshold: -0.50%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	9	3	12
No Crash	566	571	1137
Total	575	574	1149

Table H.2.3.6: Out-of-Sample Market Heat LDA Confusion Matrices for BG Group Stock (December 2007)

Crash Threshold: -0.10%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	27	52	79
No Crash	137	428	565
Total	164	480	644

Crash Threshold: -0.25%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	10	20	30
No Crash	144	470	614
Total	154	490	644

Crash Threshold: -0.50%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	0	2	2
No Crash	116	526	642
Total	116	528	644

Table H.2.3.7: Out-of-Sample Market Heat LDA Confusion Matrices for BG Group Stock (January 2008)

Crash Threshold: -0.10%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	202	167	369
No Crash	441	346	787
Total	643	513	1156

Crash Threshold: -0.25%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	86	135	221
No Crash	355	580	935
Total	441	715	1156

Crash Threshold: -0.50%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	24	47	71
No Crash	507	578	1085
Total	531	625	1156

Table H.2.3.8: Out-of-Sample Market Heat LDA Confusion Matrices for BG Group Stock (February 2008)

Crash Threshold: -0.10%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	110	127	237
No Crash	324	500	824
Total	434	627	1061

Crash Threshold: -0.25%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	59	66	125
No Crash	377	559	936
Total	436	625	1061

Crash Threshold: -0.50%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	4	14	18
No Crash	431	612	1043
Total	435	626	1061

Table H.2.3.9: Out-of-Sample Market Heat LDA Confusion Matrices for BG Group Stock (March 2008)

Crash Threshold: -0.10%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	70	131	201
No Crash	256	414	670
Total	326	545	871

Crash Threshold: -0.25%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	44	64	108
No Crash	275	488	763
Total	319	552	871

Crash Threshold: -0.50%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	6	10	16
No Crash	466	389	855
Total	472	399	871

Table H.2.3.10: Out-of-Sample Market Heat LDA Confusion Matrices for BG Group Stock (April 2008)

Crash Threshold: -0.10%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	124	142	266
No Crash	404	463	867
Total	528	605	1133

Crash Threshold: -0.25%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	41	54	95
No Crash	493	545	1038
Total	534	599	1133

Crash Threshold: -0.50%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	9	10	19
No Crash	539	575	1114
Total	548	585	1133

Table H.2.3.11: Out-of-Sample Market Heat LDA Confusion Matrices for BG Group Stock (May 2008)

Crash Threshold: -0.10%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	133	125	258
No Crash	373	321	694
Total	506	446	952

Crash Threshold: -0.25%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	50	60	110
No Crash	344	498	842
Total	394	558	952

Crash Threshold: -0.50%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	9	11	20
No Crash	458	474	932
Total	467	485	952

Table H.2.3.12: Out-of-Sample Market Heat LDA Confusion Matrices for BG Group Stock (June 2008)

Crash Threshold: -0.10%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	160	141	301
No Crash	273	475	748
Total	433	616	1049

Crash Threshold: -0.25%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	62	56	118
No Crash	363	568	931
Total	425	624	1049

Crash Threshold: -0.50%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	10	12	22
No Crash	462	565	1027
Total	472	577	1049

Table H.3.1.1: Out-of-Sample Linear Crash Estimator LDA Confusion Matrices for British Petroleum Stock (July 2007)

Crash Threshold: -0.10%

Actual		Predicted: Crash	Predicted: No Crash	Total
	Crash	121	42	163
	No Crash	634	346	980
	Total	755	388	1143

Crash Threshold: -0.25%

Actual		Predicted: Crash	Predicted: No Crash	Total
	Crash	39	13	52
	No Crash	501	590	1091
	Total	540	603	1143

Crash Threshold: -0.50%

Actual		Predicted: Crash	Predicted: No Crash	Total
	Crash	0	1	1
	No Crash	193	949	1142
	Total	193	950	1143

Table H.3.1.2: Out-of-Sample Linear Crash Estimator LDA Confusion Matrices for British Petroleum Stock (August 2007)

Crash Threshold: -0.10%

Actual		Predicted: Crash	Predicted: No Crash	Total
	Crash	54	86	140
	No Crash	248	761	1009
	Total	302	847	1149

Crash Threshold: -0.25%

Actual		Predicted: Crash	Predicted: No Crash	Total
	Crash	23	26	49
	No Crash	226	874	1100
	Total	249	900	1149

Crash Threshold: -0.50%

Actual		Predicted: Crash	Predicted: No Crash	Total
	Crash	3	2	5
	No Crash	117	1027	1144
	Total	120	1029	1149

Table H.3.1.3: Out-of-Sample Linear Crash Estimator LDA Confusion Matrices for British Petroleum Stock (September 2007)

Crash Threshold: -0.10%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	48	66	114
No Crash	284	552	836
Total	332	618	950

Crash Threshold: -0.25%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	17	19	36
No Crash	235	679	914
Total	252	698	950

Crash Threshold: -0.50%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	0	1	1
No Crash	148	801	949
Total	148	802	950

Table H.3.1.4: Out-of-Sample Linear Crash Estimator LDA Confusion Matrices for British Petroleum Stock (October 2007)

Crash Threshold: -0.10%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	100	72	172
No Crash	558	512	1070
Total	658	584	1242

Crash Threshold: -0.25%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	22	5	27
No Crash	512	703	1215
Total	534	708	1242

Crash Threshold: -0.50%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	2	0	2
No Crash	730	510	1240
Total	732	510	1242

Table H.3.1.5: Out-of-Sample Linear Crash Estimator LDA Confusion Matrices for British Petroleum Stock (November 2007)

Crash Threshold: -0.10%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	81	140	221
No Crash	380	546	926
Total	461	686	1147

Crash Threshold: -0.25%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	39	50	89
No Crash	396	662	1058
Total	435	712	1147

Crash Threshold: -0.50%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	0	1	1
No Crash	90	1056	1146
Total	90	1057	1147

Table H.3.1.6: Out-of-Sample Linear Crash Estimator LDA Confusion Matrices for British Petroleum Stock (December 2007)

Crash Threshold: -0.10%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	32	52	84
No Crash	168	411	579
Total	200	463	663

Crash Threshold: -0.25%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	1	6	7
No Crash	157	499	656
Total	158	505	663

Crash Threshold: -0.50%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	0	0	0
No Crash	0	663	663
Total	0	663	663

Table H.3.1.7: Out-of-Sample Linear Crash Estimator LDA Confusion Matrices for British Petroleum Stock (January 2008)

Crash Threshold: -0.10%

Actual		Predicted: Crash	Predicted: No Crash	Total
	Crash	214	120	334
	No Crash	470	363	833
	Total	684	483	1167

Crash Threshold: -0.25%

Actual		Predicted: Crash	Predicted: No Crash	Total
	Crash	148	60	208
	No Crash	523	436	959
	Total	671	496	1167

Crash Threshold: -0.50%

Actual		Predicted: Crash	Predicted: No Crash	Total
	Crash	36	8	44
	No Crash	572	551	1123
	Total	608	559	1167

Table H.3.1.8: Out-of-Sample Linear Crash Estimator LDA Confusion Matrices for British Petroleum Stock (February 2008)

Crash Threshold: -0.10%

Actual		Predicted: Crash	Predicted: No Crash	Total
	Crash	81	118	199
	No Crash	289	563	852
	Total	370	681	1051

Crash Threshold: -0.25%

Actual		Predicted: Crash	Predicted: No Crash	Total
	Crash	38	50	88
	No Crash	291	672	963
	Total	329	722	1051

Crash Threshold: -0.50%

Actual		Predicted: Crash	Predicted: No Crash	Total
	Crash	4	3	7
	No Crash	199	845	1044
	Total	203	848	1051

Table H.3.1.9: Out-of-Sample Linear Crash Estimator LDA Confusion Matrices for British Petroleum Stock (March 2008)

Crash Threshold: -0.10%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	89	79	168
No Crash	284	416	700
Total	373	495	868

Crash Threshold: -0.25%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	37	26	63
No Crash	325	480	805
Total	362	506	868

Crash Threshold: -0.50%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	2	1	3
No Crash	118	747	865
Total	120	748	868

Table H.3.1.10: Out-of-Sample Linear Crash Estimator LDA Confusion Matrices for British Petroleum Stock (April 2008)

Crash Threshold: -0.10%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	87	80	167
No Crash	471	512	983
Total	558	592	1150

Crash Threshold: -0.25%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	33	28	61
No Crash	516	573	1089
Total	549	601	1150

Crash Threshold: -0.50%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	0	4	4
No Crash	333	813	1146
Total	333	817	1150

Table H.3.1.11: Out-of-Sample Linear Crash Estimator LDA Confusion Matrices for British Petroleum Stock (May 2008)

Crash Threshold: -0.10%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	137	64	201
No Crash	426	322	748
Total	563	386	949

Crash Threshold: -0.25%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	28	12	40
No Crash	514	395	909
Total	542	407	949

Crash Threshold: -0.50%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	0	2	2
No Crash	53	894	947
Total	53	896	949

Table H.3.1.12: Out-of-Sample Linear Crash Estimator LDA Confusion Matrices for British Petroleum Stock (June 2008)

Crash Threshold: -0.10%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	149	135	284
No Crash	310	474	784
Total	459	609	1068

Crash Threshold: -0.25%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	72	34	106
No Crash	468	494	962
Total	540	528	1068

Crash Threshold: -0.50%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	5	2	7
No Crash	460	601	1061
Total	465	603	1068

Table H.3.2.1: Out-of-Sample VPIN LDA Confusion Matrices for British Petroleum Stock (July 2007)

Crash Threshold: -0.10%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	98	65	163
No Crash	556	424	980
Total	654	489	1143

Crash Threshold: -0.25%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	31	21	52
No Crash	547	544	1091
Total	578	565	1143

Crash Threshold: -0.50%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	0	1	1
No Crash	134	1008	1142
Total	134	1009	1143

Table H.3.2.2: Out-of-Sample VPIN LDA Confusion Matrices for British Petroleum Stock (August 2007)

Crash Threshold: -0.10%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	79	61	140
No Crash	494	515	1009
Total	573	576	1149

Crash Threshold: -0.25%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	30	19	49
No Crash	515	585	1100
Total	545	604	1149

Crash Threshold: -0.50%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	0	5	5
No Crash	461	683	1144
Total	461	688	1149

Table H.3.2.3: Out-of-Sample VPIN LDA Confusion Matrices for British Petroleum Stock (September 2007)

Crash Threshold: -0.10%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	59	55	114
No Crash	435	401	836
Total	494	456	950

Crash Threshold: -0.25%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	17	19	36
No Crash	455	459	914
Total	472	478	950

Crash Threshold: -0.50%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	0	1	1
No Crash	85	864	949
Total	85	865	950

Table H.3.2.4: Out-of-Sample VPIN LDA Confusion Matrices for British Petroleum Stock (October 2007)

Crash Threshold: -0.10%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	102	70	172
No Crash	583	487	1070
Total	685	557	1242

Crash Threshold: -0.25%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	16	11	27
No Crash	643	572	1215
Total	659	583	1242

Crash Threshold: -0.50%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	1	1	2
No Crash	599	641	1240
Total	600	642	1242

Table H.3.2.5: Out-of-Sample VPIN LDA Confusion Matrices for British Petroleum Stock (November 2007)

Crash Threshold: -0.10%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	109	112	221
No Crash	496	430	926
Total	605	542	1147

Crash Threshold: -0.25%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	45	44	89
No Crash	540	518	1058
Total	585	562	1147

Crash Threshold: -0.50%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	1	0	1
No Crash	455	691	1146
Total	456	691	1147

Table H.3.2.6: Out-of-Sample VPIN LDA Confusion Matrices for British Petroleum Stock (December 2007)

Crash Threshold: -0.10%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	47	37	84
No Crash	259	320	579
Total	306	357	663

Crash Threshold: -0.25%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	2	5	7
No Crash	350	306	656
Total	352	311	663

Crash Threshold: -0.50%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	0	0	0
No Crash	0	663	663
Total	0	663	663

Table H.3.2.7: Out-of-Sample VPIN LDA Confusion Matrices for British Petroleum Stock (January 2008)

Crash Threshold: -0.10%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	193	141	334
No Crash	474	356	830
Total	667	497	1164

Crash Threshold: -0.25%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	122	86	208
No Crash	529	427	956
Total	651	513	1164

Crash Threshold: -0.50%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	24	20	44
No Crash	567	553	1120
Total	591	573	1164

Table H.3.2.8: Out-of-Sample VPIN LDA Confusion Matrices for British Petroleum Stock (February 2008)

Crash Threshold: -0.10%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	112	87	199
No Crash	458	394	852
Total	570	481	1051

Crash Threshold: -0.25%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	43	45	88
No Crash	478	485	963
Total	521	530	1051

Crash Threshold: -0.50%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	3	4	7
No Crash	499	545	1044
Total	502	549	1051

Table H.3.2.9: Out-of-Sample VPIN LDA Confusion Matrices for British Petroleum Stock (March 2008)

Crash Threshold: -0.10%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	88	80	168
No Crash	362	338	700
Total	450	418	868

Crash Threshold: -0.25%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	36	27	63
No Crash	410	395	805
Total	446	422	868

Crash Threshold: -0.50%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	1	2	3
No Crash	355	510	865
Total	356	512	868

Table H.3.2.10: Out-of-Sample VPIN LDA Confusion Matrices for British Petroleum Stock (April 2008)

Crash Threshold: -0.10%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	80	87	167
No Crash	429	554	983
Total	509	641	1150

Crash Threshold: -0.25%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	26	35	61
No Crash	521	568	1089
Total	547	603	1150

Crash Threshold: -0.50%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	2	2	4
No Crash	549	597	1146
Total	551	599	1150

Table H.3.2.11: Out-of-Sample VPIN LDA Confusion Matrices for British Petroleum Stock (May 2008)

Crash Threshold: -0.10%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	125	76	201
No Crash	470	278	748
Total	595	354	949

Crash Threshold: -0.25%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	23	17	40
No Crash	543	366	909
Total	566	383	949

Crash Threshold: -0.50%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	0	2	2
No Crash	125	822	947
Total	125	824	949

Table H.3.2.12: Out-of-Sample VPIN LDA Confusion Matrices for British Petroleum Stock (June 2008)

Crash Threshold: -0.10%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	154	130	284
No Crash	417	367	784
Total	571	497	1068

Crash Threshold: -0.25%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	62	44	106
No Crash	512	450	962
Total	574	494	1068

Crash Threshold: -0.50%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	1	6	7
No Crash	517	544	1061
Total	518	550	1068

Table H.3.3.1: Out-of-Sample Market Heat LDA Confusion Matrices for British Petroleum Stock (July 2007)

Crash Threshold: -0.10%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	90	73	163
No Crash	481	499	980
Total	571	572	1143

Crash Threshold: -0.25%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	24	28	52
No Crash	406	685	1091
Total	430	713	1143

Crash Threshold: -0.50%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	0	1	1
No Crash	112	1030	1142
Total	112	1031	1143

Table H.3.3.2: Out-of-Sample Market Heat LDA Confusion Matrices for British Petroleum Stock (August 2007)

Crash Threshold: -0.10%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	59	81	140
No Crash	357	652	1009
Total	416	733	1149

Crash Threshold: -0.25%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	22	27	49
No Crash	346	754	1100
Total	368	781	1149

Crash Threshold: -0.50%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	1	4	5
No Crash	371	773	1144
Total	372	777	1149

Table H.3.3.3: Out-of-Sample Market Heat LDA Confusion Matrices for British Petroleum Stock (September 2007)

Crash Threshold: -0.10%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	56	58	114
No Crash	325	511	836
Total	381	569	950

Crash Threshold: -0.25%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	20	16	36
No Crash	476	438	914
Total	496	454	950

Crash Threshold: -0.50%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	0	1	1
No Crash	79	870	949
Total	79	871	950

Table H.3.3.4: Out-of-Sample Market Heat LDA Confusion Matrices for British Petroleum Stock (October 2007)

Crash Threshold: -0.10%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	68	104	172
No Crash	364	706	1070
Total	432	810	1242

Crash Threshold: -0.25%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	13	14	27
No Crash	423	792	1215
Total	436	806	1242

Crash Threshold: -0.50%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	0	2	2
No Crash	559	681	1240
Total	559	683	1242

Table H.3.3.5: Out-of-Sample Market Heat LDA Confusion Matrices for British Petroleum Stock (November 2007)

Crash Threshold: -0.10%

Actual		Predicted: Crash	Predicted: No Crash	Total
	Crash	89	132	221
	No Crash	403	523	926
	Total	492	655	1147

Crash Threshold: -0.25%

Actual		Predicted: Crash	Predicted: No Crash	Total
	Crash	33	56	89
	No Crash	438	620	1058
	Total	471	676	1147

Crash Threshold: -0.50%

Actual		Predicted: Crash	Predicted: No Crash	Total
	Crash	0	1	1
	No Crash	551	595	1146
	Total	551	596	1147

Table H.3.3.6: Out-of-Sample Market Heat LDA Confusion Matrices for British Petroleum Stock (December 2007)

Crash Threshold: -0.10%

Actual		Predicted: Crash	Predicted: No Crash	Total
	Crash	28	56	84
	No Crash	154	425	579
	Total	182	481	663

Crash Threshold: -0.25%

Actual		Predicted: Crash	Predicted: No Crash	Total
	Crash	3	4	7
	No Crash	279	377	656
	Total	282	381	663

Crash Threshold: -0.50%

Actual		Predicted: Crash	Predicted: No Crash	Total
	Crash	0	0	0
	No Crash	0	663	663
	Total	0	663	663

Table H.3.3.7: Out-of-Sample Market Heat LDA Confusion Matrices for British Petroleum Stock (January 2008)

Crash Threshold: -0.10%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	143	191	334
No Crash	340	490	830
Total	483	681	1164

Crash Threshold: -0.25%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	82	126	208
No Crash	380	576	956
Total	462	702	1164

Crash Threshold: -0.50%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	22	22	44
No Crash	419	701	1120
Total	441	723	1164

Table H.3.3.8: Out-of-Sample Market Heat LDA Confusion Matrices for British Petroleum Stock (February 2008)

Crash Threshold: -0.10%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	74	125	199
No Crash	308	544	852
Total	382	669	1051

Crash Threshold: -0.25%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	34	54	88
No Crash	456	507	963
Total	490	561	1051

Crash Threshold: -0.50%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	2	5	7
No Crash	533	511	1044
Total	535	516	1051

Table H.3.3.9: Out-of-Sample Market Heat LDA Confusion Matrices for British Petroleum Stock (March 2008)

Crash Threshold: -0.10%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	61	107	168
No Crash	194	506	700
Total	255	613	868

Crash Threshold: -0.25%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	24	39	63
No Crash	225	580	805
Total	249	619	868

Crash Threshold: -0.50%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	0	3	3
No Crash	102	763	865
Total	102	766	868

Table H.3.3.10: Out-of-Sample Market Heat LDA Confusion Matrices for British Petroleum Stock (April 2008)

Crash Threshold: -0.10%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	82	85	167
No Crash	446	537	983
Total	528	622	1150

Crash Threshold: -0.25%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	28	33	61
No Crash	518	571	1089
Total	546	604	1150

Crash Threshold: -0.50%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	2	2	4
No Crash	537	609	1146
Total	539	611	1150

Table H.3.3.11: Out-of-Sample Market Heat LDA Confusion Matrices for British Petroleum Stock (May 2008)

Crash Threshold: -0.10%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	91	110	201
No Crash	305	443	748
Total	396	553	949

Crash Threshold: -0.25%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	15	25	40
No Crash	360	549	909
Total	375	574	949

Crash Threshold: -0.50%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	0	2	2
No Crash	146	801	947
Total	146	803	949

Table H.3.3.12: Out-of-Sample Market Heat LDA Confusion Matrices for British Petroleum Stock (June 2008)

Crash Threshold: -0.10%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	151	133	284
No Crash	409	375	784
Total	560	508	1068

Crash Threshold: -0.25%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	49	57	106
No Crash	405	557	962
Total	454	614	1068

Crash Threshold: -0.50%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	5	2	7
No Crash	556	505	1061
Total	561	507	1068

Table H.4.1.1: Out-of-Sample Linear Crash Estimator LDA Confusion Matrices for Diageo Stock (July 2007)

Crash Threshold: -0.10%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	109	37	146
No Crash	652	301	953
Total	761	338	1099

Crash Threshold: -0.25%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	21	18	39
No Crash	469	591	1060
Total	490	609	1099

Crash Threshold: -0.50%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	0	2	2
No Crash	173	924	1097
Total	173	926	1099

Table H.4.1.2: Out-of-Sample Linear Crash Estimator LDA Confusion Matrices for Diageo Stock (August 2007)

Crash Threshold: -0.10%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	54	104	158
No Crash	248	723	971
Total	302	827	1129

Crash Threshold: -0.25%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	19	27	46
No Crash	198	885	1083
Total	217	912	1129

Crash Threshold: -0.50%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	2	2	4
No Crash	195	930	1125
Total	197	932	1129

Table H.4.1.3: Out-of-Sample Linear Crash Estimator LDA Confusion Matrices for Diageo Stock (September 2007)

Crash Threshold: -0.10%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	39	47	86
No Crash	319	524	843
Total	358	571	929

Crash Threshold: -0.25%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	8	12	20
No Crash	255	654	909
Total	263	666	929

Crash Threshold: -0.50%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	0	0	0
No Crash	24	905	929
Total	24	905	929

Table H.4.1.4: Out-of-Sample Linear Crash Estimator LDA Confusion Matrices for Diageo Stock (October 2007)

Crash Threshold: -0.10%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	62	74	136
No Crash	422	645	1067
Total	484	719	1203

Crash Threshold: -0.25%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	16	30	46
No Crash	353	804	1157
Total	369	834	1203

Crash Threshold: -0.50%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	2	4	6
No Crash	200	997	1197
Total	202	1001	1203

Table H.4.1.5: Out-of-Sample Linear Crash Estimator LDA Confusion Matrices for Diageo Stock (November 2007)

Crash Threshold: -0.10%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	93	98	191
No Crash	436	507	943
Total	529	605	1134

Crash Threshold: -0.25%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	45	37	82
No Crash	591	461	1052
Total	636	498	1134

Crash Threshold: -0.50%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	3	0	3
No Crash	337	794	1131
Total	340	794	1134

Table H.4.1.6: Out-of-Sample Linear Crash Estimator LDA Confusion Matrices for Diageo Stock (December 2007)

Crash Threshold: -0.10%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	16	49	65
No Crash	124	434	558
Total	140	483	623

Crash Threshold: -0.25%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	3	15	18
No Crash	108	497	605
Total	111	512	623

Crash Threshold: -0.50%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	0	0	0
No Crash	196	427	623
Total	196	427	623

Table H.4.1.7: Out-of-Sample Linear Crash Estimator LDA Confusion Matrices for Diageo Stock (January 2008)

Crash Threshold: -0.10%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	246	156	402
No Crash	436	305	741
Total	682	461	1143

Crash Threshold: -0.25%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	131	62	193
No Crash	524	426	950
Total	655	488	1143

Crash Threshold: -0.50%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	31	16	47
No Crash	526	570	1096
Total	557	586	1143

Table H.4.1.8: Out-of-Sample Linear Crash Estimator LDA Confusion Matrices for Diageo Stock (February 2008)

Crash Threshold: -0.10%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	54	143	197
No Crash	188	672	860
Total	242	815	1057

Crash Threshold: -0.25%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	24	57	81
No Crash	181	795	976
Total	205	852	1057

Crash Threshold: -0.50%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	0	6	6
No Crash	148	903	1051
Total	148	909	1057

Table H.4.1.9: Out-of-Sample Linear Crash Estimator LDA Confusion Matrices for Diageo Stock (March 2008)

Crash Threshold: -0.10%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	42	94	136
No Crash	181	541	722
Total	223	635	858

Crash Threshold: -0.25%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	16	35	51
No Crash	188	619	807
Total	204	654	858

Crash Threshold: -0.50%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	0	2	2
No Crash	208	648	856
Total	208	650	858

Table H.4.1.10: Out-of-Sample Linear Crash Estimator LDA Confusion Matrices for Diageo Stock (April 2008)

Crash Threshold: -0.10%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	85	72	157
No Crash	442	499	941
Total	527	571	1098

Crash Threshold: -0.25%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	35	20	55
No Crash	447	596	1043
Total	482	616	1098

Crash Threshold: -0.50%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	1	0	1
No Crash	343	754	1097
Total	344	754	1098

Table H.4.1.11: Out-of-Sample Linear Crash Estimator LDA Confusion Matrices for Diageo Stock (May 2008)

Crash Threshold: -0.10%

Actual \ Predicted	Crash	No Crash	Total
Crash	165	43	208
No Crash	471	258	729
Total	636	301	937

Crash Threshold: -0.25%

Actual \ Predicted	Crash	No Crash	Total
Crash	35	11	46
No Crash	433	458	891
Total	468	469	937

Crash Threshold: -0.50%

Actual \ Predicted	Crash	No Crash	Total
Crash	3	1	4
No Crash	551	382	933
Total	554	383	937

Table H.4.1.12: Out-of-Sample Linear Crash Estimator LDA Confusion Matrices for Diageo Stock (June 2008)

Crash Threshold: -0.10%

Actual \ Predicted	Crash	No Crash	Total
Crash	123	155	278
No Crash	261	509	770
Total	384	664	1048

Crash Threshold: -0.25%

Actual \ Predicted	Crash	No Crash	Total
Crash	26	38	64
No Crash	248	736	984
Total	274	774	1048

Crash Threshold: -0.50%

Actual \ Predicted	Crash	No Crash	Total
Crash	0	0	0
No Crash	99	949	1048
Total	99	949	1048

Table H.4.2.1: Out-of-Sample VPIN LDA Confusion Matrices for Diageo Stock (July 2007)

Crash Threshold: -0.10%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	83	63	146
No Crash	468	485	953
Total	551	548	1099

Crash Threshold: -0.25%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	20	19	39
No Crash	523	537	1060
Total	543	556	1099

Crash Threshold: -0.50%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	0	2	2
No Crash	142	955	1097
Total	142	957	1099

Table H.4.2.2: Out-of-Sample VPIN LDA Confusion Matrices for Diageo Stock (August 2007)

Crash Threshold: -0.10%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	82	76	158
No Crash	457	514	971
Total	539	590	1129

Crash Threshold: -0.25%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	26	20	46
No Crash	531	552	1083
Total	557	572	1129

Crash Threshold: -0.50%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	3	1	4
No Crash	553	572	1125
Total	556	573	1129

Table H.4.2.3: Out-of-Sample VPIN LDA Confusion Matrices for Diageo Stock (September 2007)

Crash Threshold: -0.10%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	44	42	86
No Crash	425	418	843
Total	469	460	929

Crash Threshold: -0.25%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	10	10	20
No Crash	429	480	909
Total	439	490	929

Crash Threshold: -0.50%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	0	0	0
No Crash	80	849	929
Total	80	849	929

Table H.4.2.4: Out-of-Sample VPIN LDA Confusion Matrices for Diageo Stock (October 2007)

Crash Threshold: -0.10%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	73	63	136
No Crash	496	571	1067
Total	569	634	1203

Crash Threshold: -0.25%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	26	20	46
No Crash	577	580	1157
Total	603	600	1203

Crash Threshold: -0.50%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	4	2	6
No Crash	451	746	1197
Total	455	748	1203

Table H.4.2.5: Out-of-Sample VPIN LDA Confusion Matrices for Diageo Stock (November 2007)

Crash Threshold: -0.10%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	85	106	191
No Crash	471	472	943
Total	556	578	1134

Crash Threshold: -0.25%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	39	43	82
No Crash	503	549	1052
Total	542	592	1134

Crash Threshold: -0.50%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	1	2	3
No Crash	490	641	1131
Total	491	643	1134

Table H.4.2.6: Out-of-Sample VPIN LDA Confusion Matrices for Diageo Stock (December 2007)

Crash Threshold: -0.10%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	37	28	65
No Crash	309	249	558
Total	346	277	623

Crash Threshold: -0.25%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	12	6	18
No Crash	315	290	605
Total	327	296	623

Crash Threshold: -0.50%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	0	0	0
No Crash	258	365	623
Total	258	365	623

Table H.4.2.7: Out-of-Sample VPIN LDA Confusion Matrices for Diageo Stock (January 2008)

Crash Threshold: -0.10%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	261	141	402
No Crash	456	285	741
Total	717	426	1143

Crash Threshold: -0.25%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	132	61	193
No Crash	556	394	950
Total	688	455	1143

Crash Threshold: -0.50%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	31	16	47
No Crash	657	439	1096
Total	688	455	1143

Table H.4.2.8: Out-of-Sample VPIN LDA Confusion Matrices for Diageo Stock (February 2008)

Crash Threshold: -0.10%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	106	91	197
No Crash	413	447	860
Total	519	538	1057

Crash Threshold: -0.25%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	36	45	81
No Crash	445	531	976
Total	481	576	1057

Crash Threshold: -0.50%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	2	4	6
No Crash	460	591	1051
Total	462	595	1057

Table H.4.2.9: Out-of-Sample VPIN LDA Confusion Matrices for Diageo Stock (March 2008)

Crash Threshold: -0.10%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	69	67	136
No Crash	376	346	722
Total	445	413	858

Crash Threshold: -0.25%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	30	21	51
No Crash	406	401	807
Total	436	422	858

Crash Threshold: -0.50%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	2	0	2
No Crash	366	490	856
Total	368	490	858

Table H.4.2.10: Out-of-Sample VPIN LDA Confusion Matrices for Diageo Stock (April 2008)

Crash Threshold: -0.10%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	65	92	157
No Crash	420	521	941
Total	485	613	1098

Crash Threshold: -0.25%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	27	28	55
No Crash	560	483	1043
Total	587	511	1098

Crash Threshold: -0.50%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	0	1	1
No Crash	560	537	1097
Total	560	538	1098

Table H.4.2.11: Out-of-Sample VPIN LDA Confusion Matrices for Diageo Stock (May 2008)

Crash Threshold: -0.10%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	133	75	208
No Crash	403	326	729
Total	536	401	937

Crash Threshold: -0.25%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	11	35	46
No Crash	419	472	891
Total	430	507	937

Crash Threshold: -0.50%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	2	2	4
No Crash	441	492	933
Total	443	494	937

Table H.4.2.12: Out-of-Sample VPIN LDA Confusion Matrices for Diageo Stock (June 2008)

Crash Threshold: -0.10%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	132	146	278
No Crash	389	381	770
Total	521	527	1048

Crash Threshold: -0.25%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	35	29	64
No Crash	536	448	984
Total	571	477	1048

Crash Threshold: -0.50%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	0	0	0
No Crash	511	537	1048
Total	511	537	1048

Table H.4.3.1: Out-of-Sample Market Heat LDA Confusion Matrices for Diageo Stock (July 2007)

Crash Threshold: -0.10%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	66	80	146
No Crash	353	600	953
Total	419	680	1099

Crash Threshold: -0.25%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	14	25	39
No Crash	443	617	1060
Total	457	642	1099

Crash Threshold: -0.50%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	0	2	2
No Crash	150	947	1097
Total	150	949	1099

Table H.4.3.2: Out-of-Sample Market Heat LDA Confusion Matrices for Diageo Stock (August 2007)

Crash Threshold: -0.10%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	56	102	158
No Crash	278	693	971
Total	334	795	1129

Crash Threshold: -0.25%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	17	29	46
No Crash	304	779	1083
Total	321	808	1129

Crash Threshold: -0.50%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	1	3	4
No Crash	471	654	1125
Total	472	657	1129

Table H.4.3.3: Out-of-Sample Market Heat LDA Confusion Matrices for Diageo Stock (September 2007)

Crash Threshold: -0.10%

Actual		Predicted: Crash	Predicted: No Crash	Total
	Crash	42	44	86
	No Crash	321	522	843
	Total	363	566	929

Crash Threshold: -0.25%

Actual		Predicted: Crash	Predicted: No Crash	Total
	Crash	11	9	20
	No Crash	292	617	909
	Total	303	626	929

Crash Threshold: -0.50%

Actual		Predicted: Crash	Predicted: No Crash	Total
	Crash	0	0	0
	No Crash	110	819	929
	Total	110	819	929

Table H.4.3.4: Out-of-Sample Market Heat LDA Confusion Matrices for Diageo Stock (October 2007)

Crash Threshold: -0.10%

Actual		Predicted: Crash	Predicted: No Crash	Total
	Crash	59	77	136
	No Crash	454	613	1067
	Total	513	690	1203

Crash Threshold: -0.25%

Actual		Predicted: Crash	Predicted: No Crash	Total
	Crash	28	18	46
	No Crash	548	609	1157
	Total	576	627	1203

Crash Threshold: -0.50%

Actual		Predicted: Crash	Predicted: No Crash	Total
	Crash	4	2	6
	No Crash	480	717	1197
	Total	484	719	1203

Table H.4.3.5: Out-of-Sample Market Heat LDA Confusion Matrices for Diageo Stock (November 2007)

Crash Threshold: -0.10%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	82	109	191
No Crash	377	566	943
Total	459	675	1134

Crash Threshold: -0.25%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	35	47	82
No Crash	490	562	1052
Total	525	609	1134

Crash Threshold: -0.50%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	1	2	3
No Crash	356	775	1131
Total	357	777	1134

Table H.4.3.6: Out-of-Sample Market Heat LDA Confusion Matrices for Diageo Stock (December 2007)

Crash Threshold: -0.10%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	43	22	65
No Crash	360	198	558
Total	403	220	623

Crash Threshold: -0.25%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	14	4	18
No Crash	389	216	605
Total	403	220	623

Crash Threshold: -0.50%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	0	0	0
No Crash	338	285	623
Total	338	285	623

Table H.4.3.7: Out-of-Sample Market Heat LDA Confusion Matrices for Diageo Stock (January 2008)

Crash Threshold: -0.10%

Actual		Predicted: Crash	Predicted: No Crash	Total
	Crash	216	186	402
	No Crash	360	381	741
	Total	576	567	1143

Crash Threshold: -0.25%

Actual		Predicted: Crash	Predicted: No Crash	Total
	Crash	113	80	193
	No Crash	440	510	950
	Total	553	590	1143

Crash Threshold: -0.50%

Actual		Predicted: Crash	Predicted: No Crash	Total
	Crash	19	28	47
	No Crash	509	587	1096
	Total	528	615	1143

Table H.4.3.8: Out-of-Sample Market Heat LDA Confusion Matrices for Diageo Stock (February 2008)

Crash Threshold: -0.10%

Actual		Predicted: Crash	Predicted: No Crash	Total
	Crash	69	128	197
	No Crash	386	474	860
	Total	455	602	1057

Crash Threshold: -0.25%

Actual		Predicted: Crash	Predicted: No Crash	Total
	Crash	39	42	81
	No Crash	424	552	976
	Total	463	594	1057

Crash Threshold: -0.50%

Actual		Predicted: Crash	Predicted: No Crash	Total
	Crash	0	6	6
	No Crash	474	577	1051
	Total	474	583	1057

Table H.4.3.9: Out-of-Sample Market Heat LDA Confusion Matrices for Diageo Stock (March 2008)

Crash Threshold: -0.10%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	38	98	136
No Crash	213	509	722
Total	251	607	858

Crash Threshold: -0.25%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	13	38	51
No Crash	229	578	807
Total	242	616	858

Crash Threshold: -0.50%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	1	1	2
No Crash	402	454	856
Total	403	455	858

Table H.4.3.10: Out-of-Sample Market Heat LDA Confusion Matrices for Diageo Stock (April 2008)

Crash Threshold: -0.10%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	88	69	157
No Crash	536	405	941
Total	624	474	1098

Crash Threshold: -0.25%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	25	30	55
No Crash	464	579	1043
Total	489	609	1098

Crash Threshold: -0.50%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	1	0	1
No Crash	497	600	1097
Total	498	600	1098

Table H.4.3.11: Out-of-Sample Market Heat LDA Confusion Matrices for Diageo Stock (May 2008)

Crash Threshold: -0.10%

Actual		Predicted: Crash	Predicted: No Crash	Total
	Crash	108	100	208
	No Crash	297	432	729
	Total	405	532	937

Crash Threshold: -0.25%

Actual		Predicted: Crash	Predicted: No Crash	Total
	Crash	25	21	46
	No Crash	365	526	891
	Total	390	547	937

Crash Threshold: -0.50%

Actual		Predicted: Crash	Predicted: No Crash	Total
	Crash	1	3	4
	No Crash	395	538	933
	Total	396	541	937

Table H.4.3.12: Out-of-Sample Market Heat LDA Confusion Matrices for Diageo Stock (June 2008)

Crash Threshold: -0.10%

Actual		Predicted: Crash	Predicted: No Crash	Total
	Crash	110	168	278
	No Crash	303	467	770
	Total	413	635	1048

Crash Threshold: -0.25%

Actual		Predicted: Crash	Predicted: No Crash	Total
	Crash	28	36	64
	No Crash	375	609	984
	Total	403	645	1048

Crash Threshold: -0.50%

Actual		Predicted: Crash	Predicted: No Crash	Total
	Crash	0	0	0
	No Crash	527	521	1048
	Total	527	521	1048

Table H.5.1.1: Out-of-Sample Linear Crash Estimator LDA Confusion Matrices for GlaxoSmithKline Stock (July 2007)

Crash Threshold: -0.10%

Actual		**Predicted** Crash	No Crash	Total
	Crash	140	56	196
	No Crash	607	323	930
	Total	747	379	1126

Crash Threshold: -0.25%

Actual		**Predicted** Crash	No Crash	Total
	Crash	27	13	40
	No Crash	671	415	1086
	Total	698	428	1126

Crash Threshold: -0.50%

Actual		**Predicted** Crash	No Crash	Total
	Crash	0	4	4
	No Crash	378	744	1122
	Total	378	748	1126

Table H.5.1.2: Out-of-Sample Linear Crash Estimator LDA Confusion Matrices for GlaxoSmithKline Stock (August 2007)

Crash Threshold: -0.10%

Actual		**Predicted** Crash	No Crash	Total
	Crash	48	111	159
	No Crash	227	764	991
	Total	275	875	1150

Crash Threshold: -0.25%

Actual		**Predicted** Crash	No Crash	Total
	Crash	15	12	27
	No Crash	192	931	1123
	Total	207	943	1150

Crash Threshold: -0.50%

Actual		**Predicted** Crash	No Crash	Total
	Crash	3	2	5
	No Crash	155	990	1145
	Total	158	992	1150

Table H.5.1.3: Out-of-Sample Linear Crash Estimator LDA Confusion Matrices for GlaxoSmithKline Stock (September 2007)

Crash Threshold: -0.10%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	65	82	147
No Crash	331	462	793
Total	396	544	940

Crash Threshold: -0.25%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	8	8	16
No Crash	344	580	924
Total	352	588	940

Crash Threshold: -0.50%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	0	0	0
No Crash	4	936	940
Total	4	936	940

Table H.5.1.4: Out-of-Sample Linear Crash Estimator LDA Confusion Matrices for GlaxoSmithKline Stock (October 2007)

Crash Threshold: -0.10%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	104	80	184
No Crash	449	606	1055
Total	553	686	1239

Crash Threshold: -0.25%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	11	12	23
No Crash	408	808	1216
Total	419	820	1239

Crash Threshold: -0.50%

Actual	Crash	No Crash	Total
Crash	1	1	2
No Crash	82	1155	1237
Total	83	1156	1239

Table H.5.1.5: Out-of-Sample Linear Crash Estimator LDA Confusion Matrices for GlaxoSmithKline Stock (November 2007)

Crash Threshold: -0.10%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	112	97	209
No Crash	493	443	936
Total	605	540	1145

Crash Threshold: -0.25%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	28	38	66
No Crash	425	654	1079
Total	453	692	1145

Crash Threshold: -0.50%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	6	1	7
No Crash	302	836	1138
Total	308	837	1145

Table H.5.1.6: Out-of-Sample Linear Crash Estimator LDA Confusion Matrices for GlaxoSmithKline Stock (December 2007)

Crash Threshold: -0.10%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	18	78	96
No Crash	112	449	561
Total	130	527	657

Crash Threshold: -0.25%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	1	14	15
No Crash	87	555	642
Total	88	569	657

Crash Threshold: -0.50%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	0	1	1
No Crash	238	418	656
Total	238	419	657

Table H.5.1.7: Out-of-Sample Linear Crash Estimator LDA Confusion Matrices for GlaxoSmithKline Stock (January 2008)

Crash Threshold: -0.10%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	196	140	336
No Crash	417	397	814
Total	613	537	1150

Crash Threshold: -0.25%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	99	68	167
No Crash	473	510	983
Total	572	578	1150

Crash Threshold: -0.50%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	28	13	41
No Crash	520	589	1109
Total	548	602	1150

Table H.5.1.8: Out-of-Sample Linear Crash Estimator LDA Confusion Matrices for GlaxoSmithKline Stock (February 2008)

Crash Threshold: -0.10%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	73	109	182
No Crash	291	589	880
Total	364	698	1062

Crash Threshold: -0.25%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	29	48	77
No Crash	356	629	985
Total	385	677	1062

Crash Threshold: -0.50%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	1	4	5
No Crash	267	790	1057
Total	268	794	1062

Table H.5.1.9: Out-of-Sample Linear Crash Estimator LDA Confusion Matrices for GlaxoSmithKline Stock (March 2008)

Crash Threshold: -0.10%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	83	82	165
No Crash	342	349	691
Total	425	431	856

Crash Threshold: -0.25%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	29	46	75
No Crash	337	444	781
Total	366	490	856

Crash Threshold: -0.50%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	2	4	6
No Crash	339	511	850
Total	341	515	856

Table H.5.1.10: Out-of-Sample Linear Crash Estimator LDA Confusion Matrices for GlaxoSmithKline Stock (April 2008)

Crash Threshold: -0.10%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	108	100	208
No Crash	417	485	902
Total	525	585	1110

Crash Threshold: -0.25%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	37	29	66
No Crash	421	623	1044
Total	458	652	1110

Crash Threshold: -0.50%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	1	4	5
No Crash	130	975	1105
Total	131	979	1110

Table H.5.1.11: Out-of-Sample Linear Crash Estimator LDA Confusion Matrices for GlaxoSmithKline Stock (May 2008)

Crash Threshold: -0.10%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	141	51	192
No Crash	459	251	710
Total	600	302	902

Crash Threshold: -0.25%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	29	17	46
No Crash	428	428	856
Total	457	445	902

Crash Threshold: -0.50%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	2	1	3
No Crash	390	509	899
Total	392	510	902

Table H.5.1.12: Out-of-Sample Linear Crash Estimator LDA Confusion Matrices for GlaxoSmithKline Stock (June 2008)

Crash Threshold: -0.10%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	150	114	264
No Crash	344	451	795
Total	494	565	1059

Crash Threshold: -0.25%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	40	37	77
No Crash	372	610	982
Total	412	647	1059

Crash Threshold: -0.50%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	5	6	11
No Crash	384	664	1048
Total	389	670	1059

Table H.5.2.1: Out-of-Sample VPIN LDA Confusion Matrices for GlaxoSmithKline Stock (July 2007)

Crash Threshold: -0.10%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	129	67	196
No Crash	569	361	930
Total	698	428	1126

Crash Threshold: -0.25%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	23	17	40
No Crash	609	477	1086
Total	632	494	1126

Crash Threshold: -0.50%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	0	4	4
No Crash	308	814	1122
Total	308	818	1126

Table H.5.2.2: Out-of-Sample VPIN LDA Confusion Matrices for GlaxoSmithKline Stock (August 2007)

Crash Threshold: -0.10%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	81	78	159
No Crash	531	460	991
Total	612	538	1150

Crash Threshold: -0.25%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	20	7	27
No Crash	538	585	1123
Total	558	592	1150

Crash Threshold: -0.50%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	4	1	5
No Crash	554	591	1145
Total	558	592	1150

Table H.5.2.3: Out-of-Sample VPIN LDA Confusion Matrices for GlaxoSmithKline Stock (September 2007)

Crash Threshold: -0.10%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	79	68	147
No Crash	401	392	793
Total	480	460	940

Crash Threshold: -0.25%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	7	9	16
No Crash	479	445	924
Total	486	454	940

Crash Threshold: -0.50%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	0	0	0
No Crash	105	835	940
Total	105	835	940

Table H.5.2.4: Out-of-Sample VPIN LDA Confusion Matrices for GlaxoSmithKline Stock (October 2007)

Crash Threshold: -0.10%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	88	96	184
No Crash	505	550	1055
Total	593	646	1239

Crash Threshold: -0.25%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	9	14	23
No Crash	516	700	1216
Total	525	714	1239

Crash Threshold: -0.50%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	1	1	2
No Crash	244	993	1237
Total	245	994	1239

Table H.5.2.5: Out-of-Sample VPIN LDA Confusion Matrices for GlaxoSmithKline Stock (November 2007)

Crash Threshold: -0.10%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	121	88	209
No Crash	510	426	936
Total	631	514	1145

Crash Threshold: -0.25%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	34	32	66
No Crash	510	569	1079
Total	544	601	1145

Crash Threshold: -0.50%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	7	0	7
No Crash	503	635	1138
Total	510	635	1145

Table H.5.2.6: Out-of-Sample VPIN LDA Confusion Matrices for GlaxoSmithKline Stock (December 2007)

Crash Threshold: -0.10%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	45	51	96
No Crash	271	290	561
Total	316	341	657

Crash Threshold: -0.25%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	7	8	15
No Crash	326	316	642
Total	333	324	657

Crash Threshold: -0.50%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	0	1	1
No Crash	279	377	656
Total	279	378	657

Table H.5.2.7: Out-of-Sample VPIN LDA Confusion Matrices for GlaxoSmithKline Stock (January 2008)

Crash Threshold: -0.10%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	206	130	336
No Crash	474	340	814
Total	680	470	1150

Crash Threshold: -0.25%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	106	61	167
No Crash	547	436	983
Total	653	497	1150

Crash Threshold: -0.50%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	20	21	41
No Crash	508	601	1109
Total	528	622	1150

Table H.5.2.8: Out-of-Sample VPIN LDA Confusion Matrices for GlaxoSmithKline Stock (February 2008)

Crash Threshold: -0.10%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	104	78	182
No Crash	452	428	880
Total	556	506	1062

Crash Threshold: -0.25%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	41	36	77
No Crash	505	480	985
Total	546	516	1062

Crash Threshold: -0.50%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	3	2	5
No Crash	465	592	1057
Total	468	594	1062

Table H.5.2.9: Out-of-Sample VPIN LDA Confusion Matrices for GlaxoSmithKline Stock (March 2008)

Crash Threshold: -0.10%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	83	82	165
No Crash	389	302	691
Total	472	384	856

Crash Threshold: -0.25%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	31	44	75
No Crash	436	345	781
Total	467	389	856

Crash Threshold: -0.50%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	1	5	6
No Crash	385	465	850
Total	386	470	856

Table H.5.2.10: Out-of-Sample VPIN LDA Confusion Matrices for GlaxoSmithKline Stock (April 2008)

Crash Threshold: -0.10%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	118	90	208
No Crash	455	447	902
Total	573	537	1110

Crash Threshold: -0.25%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	39	27	66
No Crash	533	511	1044
Total	572	538	1110

Crash Threshold: -0.50%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	4	1	5
No Crash	416	689	1105
Total	420	690	1110

Table H.5.2.11: Out-of-Sample VPIN LDA Confusion Matrices for GlaxoSmithKline Stock (May 2008)

Crash Threshold: -0.10%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	132	60	192
No Crash	432	278	710
Total	564	338	902

Crash Threshold: -0.25%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	32	14	46
No Crash	514	342	856
Total	546	356	902

Crash Threshold: -0.50%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	3	0	3
No Crash	395	504	899
Total	398	504	902

Table H.5.2.12: Out-of-Sample VPIN LDA Confusion Matrices for GlaxoSmithKline Stock (June 2008)

Crash Threshold: -0.10%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	153	111	264
No Crash	471	324	795
Total	624	435	1059

Crash Threshold: -0.25%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	45	32	77
No Crash	517	465	982
Total	562	497	1059

Crash Threshold: -0.50%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	8	3	11
No Crash	538	510	1048
Total	546	513	1059

Table H.5.3.1: Out-of-Sample Market Heat LDA Confusion Matrices for GlaxoSmithKline Stock (July 2007)

Crash Threshold: -0.10%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	89	107	196
No Crash	362	568	930
Total	451	675	1126

Crash Threshold: -0.25%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	22	18	40
No Crash	471	615	1086
Total	493	633	1126

Crash Threshold: -0.50%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	1	3	4
No Crash	275	847	1122
Total	276	850	1126

Table H.5.3.2: Out-of-Sample Market Heat LDA Confusion Matrices for GlaxoSmithKline Stock (August 2007)

Crash Threshold: -0.10%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	69	90	159
No Crash	342	649	991
Total	411	739	1150

Crash Threshold: -0.25%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	14	13	27
No Crash	335	788	1123
Total	349	801	1150

Crash Threshold: -0.50%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	1	4	5
No Crash	322	823	1145
Total	323	827	1150

Table H.5.3.3: Out-of-Sample Market Heat LDA Confusion Matrices for GlaxoSmithKline Stock (September 2007)

Crash Threshold: -0.10%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	52	95	147
No Crash	299	494	793
Total	351	589	940

Crash Threshold: -0.25%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	5	11	16
No Crash	311	613	924
Total	316	624	940

Crash Threshold: -0.50%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	0	0	0
No Crash	105	835	940
Total	105	835	940

Table H.5.3.4: Out-of-Sample Market Heat LDA Confusion Matrices for GlaxoSmithKline Stock (October 2007)

Crash Threshold: -0.10%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	84	100	184
No Crash	418	637	1055
Total	502	737	1239

Crash Threshold: -0.25%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	11	12	23
No Crash	659	557	1216
Total	670	569	1239

Crash Threshold: -0.50%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	1	1	2
No Crash	170	1067	1237
Total	171	1068	1239

Table H.5.3.5: Out-of-Sample Market Heat LDA Confusion Matrices for GlaxoSmithKline Stock (November 2007)

Crash Threshold: -0.10%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	105	104	209
No Crash	468	468	936
Total	573	572	1145

Crash Threshold: -0.25%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	29	37	66
No Crash	505	574	1079
Total	534	611	1145

Crash Threshold: -0.50%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	6	1	7
No Crash	388	750	1138
Total	394	751	1145

Table H.5.3.6: Out-of-Sample Market Heat LDA Confusion Matrices for GlaxoSmithKline Stock (December 2007)

Crash Threshold: -0.10%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	31	65	96
No Crash	156	405	561
Total	187	470	657

Crash Threshold: -0.25%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	4	11	15
No Crash	165	477	642
Total	169	488	657

Crash Threshold: -0.50%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	0	1	1
No Crash	150	506	656
Total	150	507	657

Table H.5.3.7: Out-of-Sample Market Heat LDA Confusion Matrices for GlaxoSmithKline Stock (January 2008)

Crash Threshold: -0.10%

Actual		Predicted: Crash	Predicted: No Crash	Total
	Crash	155	181	336
	No Crash	339	475	814
	Total	494	656	1150

Crash Threshold: -0.25%

Actual		Predicted: Crash	Predicted: No Crash	Total
	Crash	82	85	167
	No Crash	380	603	983
	Total	462	688	1150

Crash Threshold: -0.50%

Actual		Predicted: Crash	Predicted: No Crash	Total
	Crash	16	25	41
	No Crash	435	674	1109
	Total	451	699	1150

Table H.5.3.8: Out-of-Sample Market Heat LDA Confusion Matrices for GlaxoSmithKline Stock (February 2008)

Crash Threshold: -0.10%

Actual		Predicted: Crash	Predicted: No Crash	Total
	Crash	76	106	182
	No Crash	358	522	880
	Total	434	628	1062

Crash Threshold: -0.25%

Actual		Predicted: Crash	Predicted: No Crash	Total
	Crash	38	39	77
	No Crash	434	551	985
	Total	472	590	1062

Crash Threshold: -0.50%

Actual		Predicted: Crash	Predicted: No Crash	Total
	Crash	1	4	5
	No Crash	451	606	1057
	Total	452	610	1062

Table H.5.3.9: Out-of-Sample Market Heat LDA Confusion Matrices for GlaxoSmithKline Stock (March 2008)

Crash Threshold: -0.10%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	82	83	165
No Crash	337	354	691
Total	419	437	856

Crash Threshold: -0.25%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	55	20	75
No Crash	489	292	781
Total	544	312	856

Crash Threshold: -0.50%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	3	3	6
No Crash	482	368	850
Total	485	371	856

Table H.5.3.10: Out-of-Sample Market Heat LDA Confusion Matrices for GlaxoSmithKline Stock (April 2008)

Crash Threshold: -0.10%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	92	116	208
No Crash	378	524	902
Total	470	640	1110

Crash Threshold: -0.25%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	25	41	66
No Crash	490	554	1044
Total	515	595	1110

Crash Threshold: -0.50%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	2	3	5
No Crash	458	647	1105
Total	460	650	1110

Table H.5.3.11: Out-of-Sample Market Heat LDA Confusion Matrices for GlaxoSmithKline Stock (May 2008)

Crash Threshold: -0.10%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	124	68	192
No Crash	404	306	710
Total	528	374	902

Crash Threshold: -0.25%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	34	12	46
No Crash	470	386	856
Total	504	398	902

Crash Threshold: -0.50%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	3	0	3
No Crash	263	636	899
Total	266	636	902

Table H.5.3.12: Out-of-Sample Market Heat LDA Confusion Matrices for GlaxoSmithKline Stock (June 2008)

Crash Threshold: -0.10%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	112	152	264
No Crash	330	465	795
Total	442	617	1059

Crash Threshold: -0.25%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	31	46	77
No Crash	399	583	982
Total	430	629	1059

Crash Threshold: -0.50%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	5	6	11
No Crash	509	539	1048
Total	514	545	1059

Table H.6.1.1: Out-of-Sample Linear Crash Estimator LDA Confusion Matrices for HSBC Stock (July 2007)

Crash Threshold: -0.10%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	114	61	175
No Crash	470	477	947
Total	584	538	1122

Crash Threshold: -0.25%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	11	5	16
No Crash	599	507	1106
Total	610	512	1122

Crash Threshold: -0.50%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	0	1	1
No Crash	16	1105	1121
Total	16	1106	1122

Table H.6.1.2: Out-of-Sample Linear Crash Estimator LDA Confusion Matrices for HSBC Stock (August 2007)

Crash Threshold: -0.10%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	69	147	216
No Crash	292	651	943
Total	361	798	1159

Crash Threshold: -0.25%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	11	10	21
No Crash	240	898	1138
Total	251	908	1159

Crash Threshold: -0.50%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	0	1	1
No Crash	59	1099	1158
Total	59	1100	1159

Table H.6.1.3: Out-of-Sample Linear Crash Estimator LDA Confusion Matrices for HSBC Stock (September 2007)

Crash Threshold: -0.10%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	71	66	137
No Crash	296	525	821
Total	367	591	958

Crash Threshold: -0.25%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	4	4	8
No Crash	305	645	950
Total	309	649	958

Crash Threshold: -0.50%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	0	0	0
No Crash	165	793	958
Total	165	793	958

Table H.6.1.4: Out-of-Sample Linear Crash Estimator LDA Confusion Matrices for HSBC Stock (October 2007)

Crash Threshold: -0.10%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	127	116	243
No Crash	411	606	1017
Total	538	722	1260

Crash Threshold: -0.25%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	15	7	22
No Crash	368	870	1238
Total	383	877	1260

Crash Threshold: -0.50%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	0	1	1
No Crash	23	1236	1259
Total	23	1237	1260

Table H.6.1.5: Out-of-Sample Linear Crash Estimator LDA Confusion Matrices for HSBC Stock (November 2007)

Crash Threshold: -0.10%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	140	166	306
No Crash	337	525	862
Total	477	691	1168

Crash Threshold: -0.25%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	30	38	68
No Crash	356	744	1100
Total	386	782	1168

Crash Threshold: -0.50%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	2	5	7
No Crash	400	761	1161
Total	402	766	1168

Table H.6.1.6: Out-of-Sample Linear Crash Estimator LDA Confusion Matrices for HSBC Stock (December 2007)

Crash Threshold: -0.10%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	31	76	107
No Crash	123	434	557
Total	154	510	664

Crash Threshold: -0.25%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	2	4	6
No Crash	95	563	658
Total	97	567	664

Crash Threshold: -0.50%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	0	1	1
No Crash	69	594	663
Total	69	595	664

Table H.6.1.7: Out-of-Sample Linear Crash Estimator LDA Confusion Matrices for HSBC Stock (January 2008)

Crash Threshold: -0.10%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	253	145	398
No Crash	459	306	765
Total	712	451	1163

Crash Threshold: -0.25%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	135	57	192
No Crash	543	428	971
Total	678	485	1163

Crash Threshold: -0.50%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	40	16	56
No Crash	446	661	1107
Total	486	677	1163

Table H.6.1.8: Out-of-Sample Linear Crash Estimator LDA Confusion Matrices for HSBC Stock (February 2008)

Crash Threshold: -0.10%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	128	146	274
No Crash	339	450	789
Total	467	596	1063

Crash Threshold: -0.25%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	34	46	80
No Crash	351	632	983
Total	385	678	1063

Crash Threshold: -0.50%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	2	1	3
No Crash	241	819	1060
Total	243	820	1063

Table H.6.1.9: Out-of-Sample Linear Crash Estimator LDA Confusion Matrices for HSBC Stock (March 2008)

Crash Threshold: -0.10%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	114	109	223
No Crash	263	391	654
Total	377	500	877

Crash Threshold: -0.25%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	41	20	61
No Crash	229	587	816
Total	270	607	877

Crash Threshold: -0.50%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	5	4	9
No Crash	139	729	868
Total	144	733	877

Table H.6.1.10: Out-of-Sample Linear Crash Estimator LDA Confusion Matrices for HSBC Stock (April 2008)

Crash Threshold: -0.10%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	112	108	220
No Crash	401	542	943
Total	513	650	1163

Crash Threshold: -0.25%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	9	9	18
No Crash	511	634	1145
Total	520	643	1163

Crash Threshold: -0.50%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	0	0	0
No Crash	148	1015	1163
Total	148	1015	1163

Table H.6.1.11: Out-of-Sample Linear Crash Estimator LDA Confusion Matrices for HSBC Stock (May 2008)

Crash Threshold: -0.10%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	82	46	128
No Crash	387	431	818
Total	469	477	946

Crash Threshold: -0.25%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	10	1	11
No Crash	459	476	935
Total	469	477	946

Crash Threshold: -0.50%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	0	0	0
No Crash	0	946	946
Total	0	946	946

Table H.6.1.12: Out-of-Sample Linear Crash Estimator LDA Confusion Matrices for HSBC Stock (June 2008)

Crash Threshold: -0.10%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	122	99	221
No Crash	346	498	844
Total	468	597	1065

Crash Threshold: -0.25%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	39	12	51
No Crash	537	477	1014
Total	576	489	1065

Crash Threshold: -0.50%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	0	2	2
No Crash	160	903	1063
Total	160	905	1065

Table H.6.2.1: Out-of-Sample VPIN LDA Confusion Matrices for HSBC Stock (July 2007)

Crash Threshold: -0.10%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	112	63	175
No Crash	536	411	947
Total	648	474	1122

Crash Threshold: -0.25%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	15	1	16
No Crash	601	505	1106
Total	616	506	1122

Crash Threshold: -0.50%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	0	1	1
No Crash	170	951	1121
Total	170	952	1122

Table H.6.2.2: Out-of-Sample VPIN LDA Confusion Matrices for HSBC Stock (August 2007)

Crash Threshold: -0.10%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	113	103	216
No Crash	481	462	943
Total	594	565	1159

Crash Threshold: -0.25%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	11	10	21
No Crash	566	572	1138
Total	577	582	1159

Crash Threshold: -0.50%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	0	1	1
No Crash	329	829	1158
Total	329	830	1159

Table H.6.2.3: Out-of-Sample VPIN LDA Confusion Matrices for HSBC Stock (September 2007)

Crash Threshold: -0.10%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	78	59	137
No Crash	427	394	821
Total	505	453	958

Crash Threshold: -0.25%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	3	5	8
No Crash	447	503	950
Total	450	508	958

Crash Threshold: -0.50%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	0	0	0
No Crash	324	634	958
Total	324	634	958

Table H.6.2.4: Out-of-Sample VPIN LDA Confusion Matrices for HSBC Stock (October 2007)

Crash Threshold: -0.10%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	136	107	243
No Crash	555	462	1017
Total	691	569	1260

Crash Threshold: -0.25%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	14	8	22
No Crash	611	627	1238
Total	625	635	1260

Crash Threshold: -0.50%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	0	1	1
No Crash	155	1104	1259
Total	155	1105	1260

Table H.6.2.5: Out-of-Sample VPIN LDA Confusion Matrices for HSBC Stock (November 2007)

Crash Threshold: -0.10%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	184	122	306
No Crash	456	406	862
Total	640	528	1168

Crash Threshold: -0.25%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	42	26	68
No Crash	582	518	1100
Total	624	544	1168

Crash Threshold: -0.50%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	6	1	7
No Crash	551	610	1161
Total	557	611	1168

Table H.6.2.6: Out-of-Sample VPIN LDA Confusion Matrices for HSBC Stock (December 2007)

Crash Threshold: -0.10%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	46	61	107
No Crash	287	270	557
Total	333	331	664

Crash Threshold: -0.25%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	3	3	6
No Crash	290	368	658
Total	293	371	664

Crash Threshold: -0.50%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	1	0	1
No Crash	237	426	663
Total	238	426	664

Table H.6.2.7: Out-of-Sample VPIN LDA Confusion Matrices for HSBC Stock (January 2008)

Crash Threshold: -0.10%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	210	188	398
No Crash	432	333	765
Total	642	521	1163

Crash Threshold: -0.25%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	96	96	192
No Crash	489	482	971
Total	585	578	1163

Crash Threshold: -0.50%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	29	27	56
No Crash	607	500	1107
Total	636	527	1163

Table H.6.2.8: Out-of-Sample VPIN LDA Confusion Matrices for HSBC Stock (February 2008)

Crash Threshold: -0.10%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	155	119	274
No Crash	423	366	789
Total	578	485	1063

Crash Threshold: -0.25%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	46	34	80
No Crash	528	455	983
Total	574	489	1063

Crash Threshold: -0.50%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	3	0	3
No Crash	495	565	1060
Total	498	565	1063

Table H.6.2.9: Out-of-Sample VPIN LDA Confusion Matrices for HSBC Stock (March 2008)

Crash Threshold: -0.10%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	131	92	223
No Crash	356	298	654
Total	487	390	877

Crash Threshold: -0.25%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	35	26	61
No Crash	436	380	816
Total	471	406	877

Crash Threshold: -0.50%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	8	1	9
No Crash	406	462	868
Total	414	463	877

Table H.6.2.10: Out-of-Sample VPIN LDA Confusion Matrices for HSBC Stock (April 2008)

Crash Threshold: -0.10%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	121	99	220
No Crash	511	432	943
Total	632	531	1163

Crash Threshold: -0.25%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	8	10	18
No Crash	520	625	1145
Total	528	635	1163

Crash Threshold: -0.50%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	0	0	0
No Crash	219	944	1163
Total	219	944	1163

Table H.6.2.11: Out-of-Sample VPIN LDA Confusion Matrices for HSBC Stock (May 2008)

Crash Threshold: -0.10%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	70	58	128
No Crash	446	372	818
Total	516	430	946

Crash Threshold: -0.25%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	6	5	11
No Crash	482	453	935
Total	488	458	946

Crash Threshold: -0.50%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	0	0	0
No Crash	0	946	946
Total	0	946	946

Table H.6.2.12: Out-of-Sample VPIN LDA Confusion Matrices for HSBC Stock (June 2008)

Crash Threshold: -0.10%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	132	89	221
No Crash	478	366	844
Total	610	455	1065

Crash Threshold: -0.25%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	30	21	51
No Crash	533	481	1014
Total	563	502	1065

Crash Threshold: -0.50%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	1	1	2
No Crash	363	700	1063
Total	364	701	1065

Table H.6.3.1: Out-of-Sample Market Heat LDA Confusion Matrices for HSBC Stock (July 2007)

Crash Threshold: -0.10%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	85	90	175
No Crash	364	583	947
Total	449	673	1122

Crash Threshold: -0.25%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	8	8	16
No Crash	410	696	1106
Total	418	704	1122

Crash Threshold: -0.50%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	0	1	1
No Crash	155	966	1121
Total	155	967	1122

Table H.6.3.2: Out-of-Sample Market Heat LDA Confusion Matrices for HSBC Stock (August 2007)

Crash Threshold: -0.10%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	74	142	216
No Crash	307	636	943
Total	381	778	1159

Crash Threshold: -0.25%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	11	10	21
No Crash	265	873	1138
Total	276	883	1159

Crash Threshold: -0.50%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	0	1	1
No Crash	140	1018	1158
Total	140	1019	1159

Table H.6.3.3: Out-of-Sample Market Heat LDA Confusion Matrices for HSBC Stock (September 2007)

Crash Threshold: -0.10%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	50	87	137
No Crash	258	563	821
Total	308	650	958

Crash Threshold: -0.25%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	2	6	8
No Crash	269	681	950
Total	271	687	958

Crash Threshold: -0.50%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	0	0	0
No Crash	116	842	958
Total	116	842	958

Table H.6.3.4: Out-of-Sample Market Heat LDA Confusion Matrices for HSBC Stock (October 2007)

Crash Threshold: -0.10%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	112	131	243
No Crash	385	632	1017
Total	497	763	1260

Crash Threshold: -0.25%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	10	12	22
No Crash	411	827	1238
Total	421	839	1260

Crash Threshold: -0.50%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	0	1	1
No Crash	59	1200	1259
Total	59	1201	1260

Table H.6.3.5: Out-of-Sample Market Heat LDA Confusion Matrices for HSBC Stock (November 2007)

Crash Threshold: -0.10%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	133	173	306
No Crash	368	494	862
Total	501	667	1168

Crash Threshold: -0.25%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	30	38	68
No Crash	530	570	1100
Total	560	608	1168

Crash Threshold: -0.50%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	4	3	7
No Crash	412	749	1161
Total	416	752	1168

Table H.6.3.6: Out-of-Sample Market Heat LDA Confusion Matrices for HSBC Stock (December 2007)

Crash Threshold: -0.10%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	31	76	107
No Crash	160	397	557
Total	191	473	664

Crash Threshold: -0.25%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	4	2	6
No Crash	136	522	658
Total	140	524	664

Crash Threshold: -0.50%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	1	0	1
No Crash	114	549	663
Total	115	549	664

Table H.6.3.7: Out-of-Sample Market Heat LDA Confusion Matrices for HSBC Stock (January 2008)

Crash Threshold: -0.10%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	181	217	398
No Crash	351	414	765
Total	532	631	1163

Crash Threshold: -0.25%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	83	109	192
No Crash	461	510	971
Total	544	619	1163

Crash Threshold: -0.50%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	27	29	56
No Crash	590	517	1107
Total	617	546	1163

Table H.6.3.8: Out-of-Sample Market Heat LDA Confusion Matrices for HSBC Stock (February 2008)

Crash Threshold: -0.10%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	125	149	274
No Crash	308	481	789
Total	433	630	1063

Crash Threshold: -0.25%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	38	42	80
No Crash	402	581	983
Total	440	623	1063

Crash Threshold: -0.50%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	2	1	3
No Crash	394	666	1060
Total	396	667	1063

Table H.6.3.9: Out-of-Sample Market Heat LDA Confusion Matrices for HSBC Stock (March 2008)

Crash Threshold: -0.10%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	98	125	223
No Crash	245	409	654
Total	343	534	877

Crash Threshold: -0.25%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	27	34	61
No Crash	296	520	816
Total	323	554	877

Crash Threshold: -0.50%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	5	4	9
No Crash	278	590	868
Total	283	594	877

Table H.6.3.10: Out-of-Sample Market Heat LDA Confusion Matrices for HSBC Stock (April 2008)

Crash Threshold: -0.10%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	76	144	220
No Crash	301	642	943
Total	377	786	1163

Crash Threshold: -0.25%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	7	11	18
No Crash	579	566	1145
Total	586	577	1163

Crash Threshold: -0.50%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	0	0	0
No Crash	325	838	1163
Total	325	838	1163

Table H.6.3.11: Out-of-Sample Market Heat LDA Confusion Matrices for HSBC Stock (May 2008)

Crash Threshold: -0.10%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	69	59	128
No Crash	407	411	818
Total	476	470	946

Crash Threshold: -0.25%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	7	4	11
No Crash	463	472	935
Total	470	476	946

Crash Threshold: -0.50%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	0	0	0
No Crash	0	946	946
Total	0	946	946

Table H.6.3.12: Out-of-Sample Market Heat LDA Confusion Matrices for HSBC Stock (June 2008)

Crash Threshold: -0.10%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	97	124	221
No Crash	302	542	844
Total	399	666	1065

Crash Threshold: -0.25%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	24	27	51
No Crash	507	507	1014
Total	531	534	1065

Crash Threshold: -0.50%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	0	2	2
No Crash	459	604	1063
Total	459	606	1065

Table H.7.1.1: Out-of-Sample Linear Crash Estimator LDA Confusion Matrices for Rio Tinto Stock (July 2007)

Crash Threshold: -0.10%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	184	126	310
No Crash	365	485	850
Total	549	611	1160

Crash Threshold: -0.25%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	57	50	107
No Crash	331	722	1053
Total	388	772	1160

Crash Threshold: -0.50%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	10	9	19
No Crash	252	889	1141
Total	262	898	1160

Table H.7.1.2: Out-of-Sample Linear Crash Estimator LDA Confusion Matrices for Rio Tinto Stock (August 2007)

Crash Threshold: -0.10%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	188	164	352
No Crash	315	500	815
Total	503	664	1167

Crash Threshold: -0.25%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	97	65	162
No Crash	369	636	1005
Total	466	701	1167

Crash Threshold: -0.50%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	27	13	40
No Crash	346	781	1127
Total	373	794	1167

Table H.7.1.3: Out-of-Sample Linear Crash Estimator LDA Confusion Matrices for Rio Tinto Stock (September 2007)

Crash Threshold: -0.10%

		Predicted		
		Crash	No Crash	Total
Actual	Crash	97	165	262
	No Crash	282	426	708
	Total	379	591	970

Crash Threshold: -0.25%

		Predicted		
		Crash	No Crash	Total
Actual	Crash	34	51	85
	No Crash	294	591	885
	Total	328	642	970

Crash Threshold: -0.50%

		Predicted		
		Crash	No Crash	Total
Actual	Crash	4	5	9
	No Crash	236	725	961
	Total	240	730	970

Table H.7.1.4: Out-of-Sample Linear Crash Estimator LDA Confusion Matrices for Rio Tinto Stock (October 2007)

Crash Threshold: -0.10%

		Predicted		
		Crash	No Crash	Total
Actual	Crash	224	127	351
	No Crash	511	402	913
	Total	735	529	1264

Crash Threshold: -0.25%

		Predicted		
		Crash	No Crash	Total
Actual	Crash	87	40	127
	No Crash	581	556	1137
	Total	668	596	1264

Crash Threshold: -0.50%

		Predicted		
		Crash	No Crash	Total
Actual	Crash	13	7	20
	No Crash	529	715	1244
	Total	542	722	1264

Table H.7.1.5: Out-of-Sample Linear Crash Estimator LDA Confusion Matrices for Rio Tinto Stock (November 2007)

Crash Threshold: -0.10%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	101	299	400
No Crash	192	576	768
Total	293	875	1168

Crash Threshold: -0.25%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	76	133	209
No Crash	249	710	959
Total	325	843	1168

Crash Threshold: -0.50%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	18	40	58
No Crash	196	914	1110
Total	214	954	1168

Table H.7.1.6: Out-of-Sample Linear Crash Estimator LDA Confusion Matrices for Rio Tinto Stock (December 2007)

Crash Threshold: -0.10%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	90	88	178
No Crash	180	318	498
Total	270	406	676

Crash Threshold: -0.25%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	21	44	65
No Crash	131	480	611
Total	152	524	676

Crash Threshold: -0.50%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	3	6	9
No Crash	124	543	667
Total	127	549	676

Table H.7.1.7: Out-of-Sample Linear Crash Estimator LDA Confusion Matrices for Rio Tinto Stock (January 2008)

Crash Threshold: -0.10%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	308	142	450
No Crash	456	260	716
Total	764	402	1166

Crash Threshold: -0.25%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	201	71	272
No Crash	535	359	894
Total	736	430	1166

Crash Threshold: -0.50%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	89	29	118
No Crash	575	473	1048
Total	664	502	1166

Table H.7.1.8: Out-of-Sample Linear Crash Estimator LDA Confusion Matrices for Rio Tinto Stock (February 2008)

Crash Threshold: -0.10%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	165	147	312
No Crash	349	409	758
Total	514	556	1070

Crash Threshold: -0.25%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	58	32	90
No Crash	412	568	980
Total	470	600	1070

Crash Threshold: -0.50%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	8	6	14
No Crash	391	665	1056
Total	399	671	1070

Table H.7.1.9: Out-of-Sample Linear Crash Estimator LDA Confusion Matrices for Rio Tinto Stock (March 2008)

Crash Threshold: -0.10%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	116	174	290
No Crash	230	361	591
Total	346	535	881

Crash Threshold: -0.25%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	61	63	124
No Crash	333	424	757
Total	394	487	881

Crash Threshold: -0.50%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	13	10	23
No Crash	319	539	858
Total	332	549	881

Table H.7.1.10: Out-of-Sample Linear Crash Estimator LDA Confusion Matrices for Rio Tinto Stock (April 2008)

Crash Threshold: -0.10%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	200	126	326
No Crash	458	383	841
Total	658	509	1167

Crash Threshold: -0.25%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	75	43	118
No Crash	489	560	1049
Total	564	603	1167

Crash Threshold: -0.50%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	16	4	20
No Crash	324	823	1147
Total	340	827	1167

Table H.7.1.11: Out-of-Sample Linear Crash Estimator LDA Confusion Matrices for Rio Tinto Stock (May 2008)

Crash Threshold: -0.10%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	179	141	320
No Crash	349	301	650
Total	528	442	970

Crash Threshold: -0.25%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	88	61	149
No Crash	398	423	821
Total	486	484	970

Crash Threshold: -0.50%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	14	11	25
No Crash	387	558	945
Total	401	569	970

Table H.7.1.12: Out-of-Sample Linear Crash Estimator LDA Confusion Matrices for Rio Tinto Stock (June 2008)

Crash Threshold: -0.10%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	180	144	324
No Crash	392	352	744
Total	572	496	1068

Crash Threshold: -0.25%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	57	60	117
No Crash	427	524	951
Total	484	584	1068

Crash Threshold: -0.50%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	15	7	22
No Crash	506	540	1046
Total	521	547	1068

Table H.7.2.1: Out-of-Sample VPIN LDA Confusion Matrices for Rio Tinto Stock (July 2007)

Crash Threshold: -0.10%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	180	130	310
No Crash	486	364	850
Total	666	494	1160

Crash Threshold: -0.25%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	66	41	107
No Crash	588	465	1053
Total	654	506	1160

Crash Threshold: -0.50%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	13	6	19
No Crash	612	529	1141
Total	625	535	1160

Table H.7.2.2: Out-of-Sample VPIN LDA Confusion Matrices for Rio Tinto Stock (August 2007)

Crash Threshold: -0.10%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	178	174	352
No Crash	441	374	815
Total	619	548	1167

Crash Threshold: -0.25%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	94	68	162
No Crash	525	480	1005
Total	619	548	1167

Crash Threshold: -0.50%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	25	15	40
No Crash	597	530	1127
Total	622	545	1167

Table H.7.2.3: Out-of-Sample VPIN LDA Confusion Matrices for Rio Tinto Stock (September 2007)

Crash Threshold: -0.10%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	134	128	262
No Crash	335	373	708
Total	469	501	970

Crash Threshold: -0.25%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	45	40	85
No Crash	421	464	885
Total	466	504	970

Crash Threshold: -0.50%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	5	4	9
No Crash	436	525	961
Total	441	529	970

Table H.7.2.4: Out-of-Sample VPIN LDA Confusion Matrices for Rio Tinto Stock (October 2007)

Crash Threshold: -0.10%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	174	177	351
No Crash	463	450	913
Total	637	627	1264

Crash Threshold: -0.25%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	73	54	127
No Crash	679	458	1137
Total	752	512	1264

Crash Threshold: -0.50%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	12	8	20
No Crash	720	524	1244
Total	732	532	1264

Table H.7.2.5: Out-of-Sample VPIN LDA Confusion Matrices for Rio Tinto Stock (November 2007)

Crash Threshold: -0.10%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	203	197	400
No Crash	389	379	768
Total	592	576	1168

Crash Threshold: -0.25%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	95	114	209
No Crash	435	524	959
Total	530	638	1168

Crash Threshold: -0.50%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	26	32	58
No Crash	480	630	1110
Total	506	662	1168

Table H.7.2.6: Out-of-Sample VPIN LDA Confusion Matrices for Rio Tinto Stock (December 2007)

Crash Threshold: -0.10%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	105	73	178
No Crash	251	247	498
Total	356	320	676

Crash Threshold: -0.25%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	39	26	65
No Crash	304	307	611
Total	343	333	676

Crash Threshold: -0.50%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	4	5	9
No Crash	269	398	667
Total	273	403	676

Table H.7.2.7: Out-of-Sample VPIN LDA Confusion Matrices for Rio Tinto Stock (January 2008)

Crash Threshold: -0.10%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	270	180	450
No Crash	433	283	716
Total	703	463	1166

Crash Threshold: -0.25%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	128	144	272
No Crash	445	449	894
Total	573	593	1166

Crash Threshold: -0.50%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	73	45	118
No Crash	610	438	1048
Total	683	483	1166

Table H.7.2.8: Out-of-Sample VPIN LDA Confusion Matrices for Rio Tinto Stock (February 2008)

Crash Threshold: -0.10%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	186	126	312
No Crash	422	336	758
Total	608	462	1070

Crash Threshold: -0.25%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	55	35	90
No Crash	526	454	980
Total	581	489	1070

Crash Threshold: -0.50%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	10	4	14
No Crash	588	468	1056
Total	598	472	1070

Table H.7.2.9: Out-of-Sample VPIN LDA Confusion Matrices for Rio Tinto Stock (March 2008)

Crash Threshold: -0.10%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	159	131	290
No Crash	339	252	591
Total	498	383	881

Crash Threshold: -0.25%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	53	71	124
No Crash	392	365	757
Total	445	436	881

Crash Threshold: -0.50%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	8	15	23
No Crash	426	432	858
Total	434	447	881

Table H.7.2.10: Out-of-Sample VPIN LDA Confusion Matrices for Rio Tinto Stock (April 2008)

Crash Threshold: -0.10%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	150	176	326
No Crash	412	429	841
Total	562	605	1167

Crash Threshold: -0.25%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	55	63	118
No Crash	548	501	1049
Total	603	564	1167

Crash Threshold: -0.50%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	7	13	20
No Crash	510	637	1147
Total	517	650	1167

Table H.7.2.11: Out-of-Sample VPIN LDA Confusion Matrices for Rio Tinto Stock (May 2008)

Crash Threshold: -0.10%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	152	168	320
No Crash	314	336	650
Total	466	504	970

Crash Threshold: -0.25%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	88	61	149
No Crash	468	353	821
Total	556	414	970

Crash Threshold: -0.50%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	9	16	25
No Crash	459	486	945
Total	468	502	970

Table H.7.2.12: Out-of-Sample VPIN LDA Confusion Matrices for Rio Tinto Stock (June 2008)

Crash Threshold: -0.10%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	160	164	324
No Crash	393	351	744
Total	553	515	1068

Crash Threshold: -0.25%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	59	58	117
No Crash	523	428	951
Total	582	486	1068

Crash Threshold: -0.50%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	11	11	22
No Crash	414	632	1046
Total	425	643	1068

Table H.7.3.1: Out-of-Sample Market Heat LDA Confusion Matrices for Rio Tinto Stock (July 2007)

Crash Threshold: -0.10%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	121	189	310
No Crash	261	589	850
Total	382	778	1160

Crash Threshold: -0.25%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	52	55	107
No Crash	292	761	1053
Total	344	816	1160

Crash Threshold: -0.50%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	11	8	19
No Crash	303	838	1141
Total	314	846	1160

Table H.7.3.2: Out-of-Sample Market Heat LDA Confusion Matrices for Rio Tinto Stock (August 2007)

Crash Threshold: -0.10%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	168	184	352
No Crash	333	482	815
Total	501	666	1167

Crash Threshold: -0.25%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	75	87	162
No Crash	409	596	1005
Total	484	683	1167

Crash Threshold: -0.50%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	22	18	40
No Crash	463	664	1127
Total	485	682	1167

Table H.7.3.3: Out-of-Sample Market Heat LDA Confusion Matrices for Rio Tinto Stock (September 2007)

Crash Threshold: -0.10%

		Predicted		
		Crash	No Crash	Total
Actual	Crash	131	131	262
	No Crash	358	350	708
	Total	489	481	970

Crash Threshold: -0.25%

		Predicted		
		Crash	No Crash	Total
Actual	Crash	37	48	85
	No Crash	373	512	885
	Total	410	560	970

Crash Threshold: -0.50%

		Predicted		
		Crash	No Crash	Total
Actual	Crash	5	4	9
	No Crash	513	448	961
	Total	518	452	970

Table H.7.3.4: Out-of-Sample Market Heat LDA Confusion Matrices for Rio Tinto Stock (October 2007)

Crash Threshold: -0.10%

		Predicted		
		Crash	No Crash	Total
Actual	Crash	201	150	351
	No Crash	501	412	913
	Total	702	562	1264

Crash Threshold: -0.25%

		Predicted		
		Crash	No Crash	Total
Actual	Crash	68	59	127
	No Crash	574	563	1137
	Total	642	622	1264

Crash Threshold: -0.50%

		Predicted		
		Crash	No Crash	Total
Actual	Crash	11	9	20
	No Crash	657	587	1244
	Total	668	596	1264

Table H.7.3.5: Out-of-Sample Market Heat LDA Confusion Matrices for Rio Tinto Stock (November 2007)

Crash Threshold: -0.10%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	166	234	400
No Crash	302	466	768
Total	468	700	1168

Crash Threshold: -0.25%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	86	123	209
No Crash	379	580	959
Total	465	703	1168

Crash Threshold: -0.50%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	23	35	58
No Crash	420	690	1110
Total	443	725	1168

Table H.7.3.6: Out-of-Sample Market Heat LDA Confusion Matrices for Rio Tinto Stock (December 2007)

Crash Threshold: -0.10%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	55	123	178
No Crash	168	330	498
Total	223	453	676

Crash Threshold: -0.25%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	18	47	65
No Crash	199	412	611
Total	217	459	676

Crash Threshold: -0.50%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	2	7	9
No Crash	247	420	667
Total	249	427	676

Table H.7.3.7: Out-of-Sample Market Heat LDA Confusion Matrices for Rio Tinto Stock (January 2008)

Crash Threshold: -0.10%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	232	218	450
No Crash	305	411	716
Total	537	629	1166

Crash Threshold: -0.25%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	157	115	272
No Crash	387	507	894
Total	544	622	1166

Crash Threshold: -0.50%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	64	54	118
No Crash	458	590	1048
Total	522	644	1166

Table H.7.3.8: Out-of-Sample Market Heat LDA Confusion Matrices for Rio Tinto Stock (February 2008)

Crash Threshold: -0.10%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	161	151	312
No Crash	358	400	758
Total	519	551	1070

Crash Threshold: -0.25%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	45	45	90
No Crash	452	528	980
Total	497	573	1070

Crash Threshold: -0.50%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	7	7	14
No Crash	540	516	1056
Total	547	523	1070

Table H.7.3.9: Out-of-Sample Market Heat LDA Confusion Matrices for Rio Tinto Stock (March 2008)

Crash Threshold: -0.10%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	118	172	290
No Crash	272	319	591
Total	390	491	881

Crash Threshold: -0.25%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	58	66	124
No Crash	423	334	757
Total	481	400	881

Crash Threshold: -0.50%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	11	12	23
No Crash	491	367	858
Total	502	379	881

Table H.7.3.10: Out-of-Sample Market Heat LDA Confusion Matrices for Rio Tinto Stock (April 2008)

Crash Threshold: -0.10%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	162	164	326
No Crash	430	411	841
Total	592	575	1167

Crash Threshold: -0.25%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	52	66	118
No Crash	509	540	1049
Total	561	606	1167

Crash Threshold: -0.50%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	9	11	20
No Crash	558	589	1147
Total	567	600	1167

Table H.7.3.11: Out-of-Sample Market Heat LDA Confusion Matrices for Rio Tinto Stock (May 2008)

Crash Threshold: -0.10%

Actual		Predicted: Crash	Predicted: No Crash	Total
	Crash	143	177	320
	No Crash	326	324	650
	Total	469	501	970

Crash Threshold: -0.25%

Actual		Predicted: Crash	Predicted: No Crash	Total
	Crash	63	86	149
	No Crash	344	477	821
	Total	407	563	970

Crash Threshold: -0.50%

Actual		Predicted: Crash	Predicted: No Crash	Total
	Crash	8	17	25
	No Crash	455	490	945
	Total	463	507	970

Table H.7.3.12: Out-of-Sample Market Heat LDA Confusion Matrices for Rio Tinto Stock (June 2008)

Crash Threshold: -0.10%

Actual		Predicted: Crash	Predicted: No Crash	Total
	Crash	162	162	324
	No Crash	367	377	744
	Total	529	539	1068

Crash Threshold: -0.25%

Actual		Predicted: Crash	Predicted: No Crash	Total
	Crash	49	68	117
	No Crash	393	558	951
	Total	442	626	1068

Crash Threshold: -0.50%

Actual		Predicted: Crash	Predicted: No Crash	Total
	Crash	13	9	22
	No Crash	488	558	1046
	Total	501	567	1068

Table H.8.1.1: Out-of-Sample Linear Crash Estimator LDA Confusion Matrices for SAB Miller Stock (July 2007)

Crash Threshold: -0.10%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	142	45	187
No Crash	587	241	828
Total	729	286	1015

Crash Threshold: -0.25%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	29	16	45
No Crash	411	559	970
Total	440	575	1015

Crash Threshold: -0.50%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	1	2	3
No Crash	308	704	1012
Total	309	706	1015

Table H.8.1.2: Out-of-Sample Linear Crash Estimator LDA Confusion Matrices for SAB Miller Stock (August 2007)

Crash Threshold: -0.10%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	96	121	217
No Crash	258	623	881
Total	354	744	1098

Crash Threshold: -0.25%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	37	29	66
No Crash	252	780	1032
Total	289	809	1098

Crash Threshold: -0.50%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	3	6	9
No Crash	234	855	1089
Total	237	861	1098

Table H.8.1.3: Out-of-Sample Linear Crash Estimator LDA Confusion Matrices for SAB Miller Stock (September 2007)

Crash Threshold: -0.10%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	90	83	173
No Crash	313	411	724
Total	403	494	897

Crash Threshold: -0.25%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	19	12	31
No Crash	432	434	866
Total	451	446	897

Crash Threshold: -0.50%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	0	2	2
No Crash	167	728	895
Total	167	730	897

Table H.8.1.4: Out-of-Sample Linear Crash Estimator LDA Confusion Matrices for SAB Miller Stock (October 2007)

Crash Threshold: -0.10%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	100	147	247
No Crash	361	599	960
Total	461	746	1207

Crash Threshold: -0.25%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	30	42	72
No Crash	311	824	1135
Total	341	866	1207

Crash Threshold: -0.50%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	1	5	6
No Crash	112	1089	1201
Total	113	1094	1207

Table H.8.1.5: Out-of-Sample Linear Crash Estimator LDA Confusion Matrices for SAB Miller Stock (November 2007)

Crash Threshold: -0.10%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	142	153	295
No Crash	398	445	843
Total	540	598	1138

Crash Threshold: -0.25%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	52	42	94
No Crash	488	556	1044
Total	540	598	1138

Crash Threshold: -0.50%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	12	4	16
No Crash	302	820	1122
Total	314	824	1138

Table H.8.1.6: Out-of-Sample Linear Crash Estimator LDA Confusion Matrices for SAB Miller Stock (December 2007)

Crash Threshold: -0.10%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	57	81	138
No Crash	189	304	493
Total	246	385	631

Crash Threshold: -0.25%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	8	25	33
No Crash	191	407	598
Total	199	432	631

Crash Threshold: -0.50%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	1	5	6
No Crash	186	439	625
Total	187	444	631

Table H.8.1.7: Out-of-Sample Linear Crash Estimator LDA Confusion Matrices for SAB Miller Stock (January 2008)

Crash Threshold: -0.10%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	239	141	380
No Crash	453	313	766
Total	692	454	1146

Crash Threshold: -0.25%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	158	100	258
No Crash	484	404	888
Total	642	504	1146

Crash Threshold: -0.50%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	44	32	76
No Crash	546	524	1070
Total	590	556	1146

Table H.8.1.8: Out-of-Sample Linear Crash Estimator LDA Confusion Matrices for SAB Miller Stock (February 2008)

Crash Threshold: -0.10%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	115	139	254
No Crash	308	463	771
Total	423	602	1025

Crash Threshold: -0.25%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	64	65	129
No Crash	353	543	896
Total	417	608	1025

Crash Threshold: -0.50%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	5	11	16
No Crash	453	556	1009
Total	458	567	1025

Table H.8.1.9: Out-of-Sample Linear Crash Estimator LDA Confusion Matrices for SAB Miller Stock (March 2008)

Crash Threshold: -0.10%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	83	99	182
No Crash	220	437	657
Total	303	536	839

Crash Threshold: -0.25%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	38	48	86
No Crash	254	499	753
Total	292	547	839

Crash Threshold: -0.50%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	3	6	9
No Crash	246	584	830
Total	249	590	839

Table H.8.1.10: Out-of-Sample Linear Crash Estimator LDA Confusion Matrices for SAB Miller Stock (April 2008)

Crash Threshold: -0.10%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	135	101	236
No Crash	477	394	871
Total	612	495	1107

Crash Threshold: -0.25%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	38	56	94
No Crash	378	635	1013
Total	416	691	1107

Crash Threshold: -0.50%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	2	5	7
No Crash	276	824	1100
Total	278	829	1107

Table H.8.1.11: Out-of-Sample Linear Crash Estimator LDA Confusion Matrices for SAB Miller Stock (May 2008)

Crash Threshold: -0.10%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	111	57	168
No Crash	356	374	730
Total	467	431	898

Crash Threshold: -0.25%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	31	19	50
No Crash	398	450	848
Total	429	469	898

Crash Threshold: -0.50%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	3	4	7
No Crash	169	722	891
Total	172	726	898

Table H.8.1.12: Out-of-Sample Linear Crash Estimator LDA Confusion Matrices for SAB Miller Stock (June 2008)

Crash Threshold: -0.10%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	124	92	216
No Crash	418	373	791
Total	542	465	1007

Crash Threshold: -0.25%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	61	48	109
No Crash	515	383	898
Total	576	431	1007

Crash Threshold: -0.50%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	5	7	12
No Crash	435	560	995
Total	440	567	1007

Table H.8.2.1: Out-of-Sample VPIN LDA Confusion Matrices for SAB Miller Stock (July 2007)

Crash Threshold: -0.10%

Actual		Predicted: Crash	Predicted: No Crash	Total
	Crash	111	76	187
	No Crash	425	403	828
	Total	536	479	1015

Crash Threshold: -0.25%

Actual		Predicted: Crash	Predicted: No Crash	Total
	Crash	26	19	45
	No Crash	508	462	970
	Total	534	481	1015

Crash Threshold: -0.50%

Actual		Predicted: Crash	Predicted: No Crash	Total
	Crash	1	2	3
	No Crash	507	505	1012
	Total	508	507	1015

Table H.8.2.2: Out-of-Sample VPIN LDA Confusion Matrices for SAB Miller Stock (August 2007)

Crash Threshold: -0.10%

Actual		Predicted: Crash	Predicted: No Crash	Total
	Crash	125	92	217
	No Crash	475	406	881
	Total	600	498	1098

Crash Threshold: -0.25%

Actual		Predicted: Crash	Predicted: No Crash	Total
	Crash	41	25	66
	No Crash	537	495	1032
	Total	578	520	1098

Crash Threshold: -0.50%

Actual		Predicted: Crash	Predicted: No Crash	Total
	Crash	5	4	9
	No Crash	540	549	1089
	Total	545	553	1098

Table H.8.2.3: Out-of-Sample VPIN LDA Confusion Matrices for SAB Miller Stock (September 2007)

Crash Threshold: -0.10%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	87	86	173
No Crash	341	383	724
Total	428	469	897

Crash Threshold: -0.25%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	12	19	31
No Crash	398	468	866
Total	410	487	897

Crash Threshold: -0.50%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	0	2	2
No Crash	388	507	895
Total	388	509	897

Table H.8.2.4: Out-of-Sample VPIN LDA Confusion Matrices for SAB Miller Stock (October 2007)

Crash Threshold: -0.10%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	107	140	247
No Crash	391	569	960
Total	498	709	1207

Crash Threshold: -0.25%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	32	40	72
No Crash	499	636	1135
Total	531	676	1207

Crash Threshold: -0.50%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	3	3	6
No Crash	535	666	1201
Total	538	669	1207

Table H.8.2.5: Out-of-Sample VPIN LDA Confusion Matrices for SAB Miller Stock (November 2007)

Crash Threshold: -0.10%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	139	156	295
No Crash	435	408	843
Total	574	564	1138

Crash Threshold: -0.25%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	49	45	94
No Crash	510	534	1044
Total	559	579	1138

Crash Threshold: -0.50%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	9	7	16
No Crash	602	520	1122
Total	611	527	1138

Table H.8.2.6: Out-of-Sample VPIN LDA Confusion Matrices for SAB Miller Stock (December 2007)

Crash Threshold: -0.10%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	61	77	138
No Crash	267	226	493
Total	328	303	631

Crash Threshold: -0.25%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	15	18	33
No Crash	287	311	598
Total	302	329	631

Crash Threshold: -0.50%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	3	3	6
No Crash	293	332	625
Total	296	335	631

Table H.8.2.7: Out-of-Sample VPIN LDA Confusion Matrices for SAB Miller Stock (January 2008)

Crash Threshold: -0.10%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	226	154	380
No Crash	444	322	766
Total	670	476	1146

Crash Threshold: -0.25%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	152	106	258
No Crash	504	384	888
Total	656	490	1146

Crash Threshold: -0.50%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	48	28	76
No Crash	589	481	1070
Total	637	509	1146

Table H.8.2.8: Out-of-Sample VPIN LDA Confusion Matrices for SAB Miller Stock (February 2008)

Crash Threshold: -0.10%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	149	105	254
No Crash	427	344	771
Total	576	449	1025

Crash Threshold: -0.25%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	59	70	129
No Crash	391	505	896
Total	450	575	1025

Crash Threshold: -0.50%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	11	5	16
No Crash	525	484	1009
Total	536	489	1025

Table H.8.2.9: Out-of-Sample VPIN LDA Confusion Matrices for SAB Miller Stock (March 2008)

Crash Threshold: -0.10%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	102	80	182
No Crash	358	299	657
Total	460	379	839

Crash Threshold: -0.25%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	48	38	86
No Crash	400	353	753
Total	448	391	839

Crash Threshold: -0.50%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	4	5	9
No Crash	408	422	830
Total	412	427	839

Table H.8.2.10: Out-of-Sample VPIN LDA Confusion Matrices for SAB Miller Stock (April 2008)

Crash Threshold: -0.10%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	128	108	236
No Crash	438	433	871
Total	566	541	1107

Crash Threshold: -0.25%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	40	54	94
No Crash	439	574	1013
Total	479	628	1107

Crash Threshold: -0.50%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	4	3	7
No Crash	525	575	1100
Total	529	578	1107

Table H.8.2.11: Out-of-Sample VPIN LDA Confusion Matrices for SAB Miller Stock (May 2008)

Crash Threshold: -0.10%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	106	62	168
No Crash	427	303	730
Total	533	365	898

Crash Threshold: -0.25%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	32	18	50
No Crash	475	373	848
Total	507	391	898

Crash Threshold: -0.50%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	4	3	7
No Crash	387	504	891
Total	391	507	898

Table H.8.2.12: Out-of-Sample VPIN LDA Confusion Matrices for SAB Miller Stock (June 2008)

Crash Threshold: -0.10%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	112	104	216
No Crash	428	363	791
Total	540	467	1007

Crash Threshold: -0.25%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	51	58	109
No Crash	438	460	898
Total	489	518	1007

Crash Threshold: -0.50%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	9	3	12
No Crash	537	458	995
Total	546	461	1007

Table H.8.3.1: Out-of-Sample Market Heat LDA Confusion Matrices for SAB Miller Stock (July 2007)

Crash Threshold: -0.10%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	81	106	187
No Crash	296	532	828
Total	377	638	1015

Crash Threshold: -0.25%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	19	26	45
No Crash	277	693	970
Total	296	719	1015

Crash Threshold: -0.50%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	1	2	3
No Crash	280	732	1012
Total	281	734	1015

Table H.8.3.2: Out-of-Sample Market Heat LDA Confusion Matrices for SAB Miller Stock (August 2007)

Crash Threshold: -0.10%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	96	121	217
No Crash	344	537	881
Total	440	658	1098

Crash Threshold: -0.25%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	34	32	66
No Crash	357	675	1032
Total	391	707	1098

Crash Threshold: -0.50%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	4	5	9
No Crash	340	749	1089
Total	344	754	1098

Table H.8.3.3: Out-of-Sample Market Heat LDA Confusion Matrices for SAB Miller Stock (September 2007)

Crash Threshold: -0.10%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	66	107	173
No Crash	342	382	724
Total	408	489	897

Crash Threshold: -0.25%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	18	13	31
No Crash	489	377	866
Total	507	390	897

Crash Threshold: -0.50%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	0	2	2
No Crash	460	435	895
Total	460	437	897

Table H.8.3.4: Out-of-Sample Market Heat LDA Confusion Matrices for SAB Miller Stock (October 2007)

Crash Threshold: -0.10%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	100	147	247
No Crash	407	553	960
Total	507	700	1207

Crash Threshold: -0.25%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	33	39	72
No Crash	396	739	1135
Total	429	778	1207

Crash Threshold: -0.50%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	1	5	6
No Crash	496	705	1201
Total	497	710	1207

Table H.8.3.5: Out-of-Sample Market Heat LDA Confusion Matrices for SAB Miller Stock (November 2007)

Crash Threshold: -0.10%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	119	176	295
No Crash	392	451	843
Total	511	627	1138

Crash Threshold: -0.25%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	50	44	94
No Crash	538	506	1044
Total	588	550	1138

Crash Threshold: -0.50%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	8	8	16
No Crash	354	768	1122
Total	362	776	1138

Table H.8.3.6: Out-of-Sample Market Heat LDA Confusion Matrices for SAB Miller Stock (December 2007)

Crash Threshold: -0.10%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	68	70	138
No Crash	231	262	493
Total	299	332	631

Crash Threshold: -0.25%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	16	17	33
No Crash	285	313	598
Total	301	330	631

Crash Threshold: -0.50%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	3	3	6
No Crash	396	229	625
Total	399	232	631

Table H.8.3.7: Out-of-Sample Market Heat LDA Confusion Matrices for SAB Miller Stock (January 2008)

Crash Threshold: -0.10%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	173	207	380
No Crash	321	445	766
Total	494	652	1146

Crash Threshold: -0.25%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	118	140	258
No Crash	422	466	888
Total	540	606	1146

Crash Threshold: -0.50%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	34	42	76
No Crash	416	654	1070
Total	450	696	1146

Table H.8.3.8: Out-of-Sample Market Heat LDA Confusion Matrices for SAB Miller Stock (February 2008)

Crash Threshold: -0.10%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	107	147	254
No Crash	299	472	771
Total	406	619	1025

Crash Threshold: -0.25%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	55	74	129
No Crash	363	533	896
Total	418	607	1025

Crash Threshold: -0.50%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	11	5	16
No Crash	366	643	1009
Total	377	648	1025

Table H.8.3.9: Out-of-Sample Market Heat LDA Confusion Matrices for SAB Miller Stock (March 2008)

Crash Threshold: -0.10%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	77	105	182
No Crash	223	434	657
Total	300	539	839

Crash Threshold: -0.25%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	30	56	86
No Crash	257	496	753
Total	287	552	839

Crash Threshold: -0.50%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	2	7	9
No Crash	284	546	830
Total	286	553	839

Table H.8.3.10: Out-of-Sample Market Heat LDA Confusion Matrices for SAB Miller Stock (April 2008)

Crash Threshold: -0.10%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	111	125	236
No Crash	399	472	871
Total	510	597	1107

Crash Threshold: -0.25%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	36	58	94
No Crash	451	562	1013
Total	487	620	1107

Crash Threshold: -0.50%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	4	3	7
No Crash	517	583	1100
Total	521	586	1107

Table H.8.3.11: Out-of-Sample Market Heat LDA Confusion Matrices for SAB Miller Stock (May 2008)

Crash Threshold: -0.10%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	80	88	168
No Crash	362	368	730
Total	442	456	898

Crash Threshold: -0.25%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	23	27	50
No Crash	350	498	848
Total	373	525	898

Crash Threshold: -0.50%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	1	6	7
No Crash	250	641	891
Total	251	647	898

Table H.8.3.12: Out-of-Sample Market Heat LDA Confusion Matrices for SAB Miller Stock (June 2008)

Crash Threshold: -0.10%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	104	112	216
No Crash	388	403	791
Total	492	515	1007

Crash Threshold: -0.25%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	67	42	109
No Crash	520	378	898
Total	587	420	1007

Crash Threshold: -0.50%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	5	7	12
No Crash	552	443	995
Total	557	450	1007

Table H.9.1.1: Out-of-Sample Linear Crash Estimator LDA Confusion Matrices for Shell Stock (July 2007)

Crash Threshold: -0.10%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	125	60	185
No Crash	486	439	925
Total	611	499	1110

Crash Threshold: -0.25%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	36	11	47
No Crash	614	449	1063
Total	650	460	1110

Crash Threshold: -0.50%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	2	3	5
No Crash	544	561	1105
Total	546	564	1110

Table H.9.1.2: Out-of-Sample Linear Crash Estimator LDA Confusion Matrices for Shell Stock (August 2007)

Crash Threshold: -0.10%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	133	126	259
No Crash	365	516	881
Total	498	642	1140

Crash Threshold: -0.25%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	25	14	39
No Crash	373	728	1101
Total	398	742	1140

Crash Threshold: -0.50%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	5	1	6
No Crash	168	966	1134
Total	173	967	1140

Table H.9.1.3: Out-of-Sample Linear Crash Estimator LDA Confusion Matrices for Shell Stock (September 2007)

Crash Threshold: -0.10%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	68	45	113
No Crash	356	464	820
Total	424	509	933

Crash Threshold: -0.25%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	5	5	10
No Crash	301	622	923
Total	306	627	933

Crash Threshold: -0.50%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	0	0	0
No Crash	0	933	933
Total	0	933	933

Table H.9.1.4: Out-of-Sample Linear Crash Estimator LDA Confusion Matrices for Shell Stock (October 2007)

Crash Threshold: -0.10%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	117	56	173
No Crash	510	552	1062
Total	627	608	1235

Crash Threshold: -0.25%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	14	9	23
No Crash	457	755	1212
Total	471	764	1235

Crash Threshold: -0.50%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	0	0	0
No Crash	143	1092	1235
Total	143	1092	1235

Table H.9.1.5: Out-of-Sample Linear Crash Estimator LDA Confusion Matrices for Shell Stock (November 2007)

Crash Threshold: -0.10%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	182	125	307
No Crash	493	336	829
Total	675	461	1136

Crash Threshold: -0.25%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	26	43	69
No Crash	365	702	1067
Total	391	745	1136

Crash Threshold: -0.50%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	1	3	4
No Crash	326	806	1132
Total	327	809	1136

Table H.9.1.6: Out-of-Sample Linear Crash Estimator LDA Confusion Matrices for Shell Stock (December 2007)

Crash Threshold: -0.10%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	39	39	78
No Crash	174	387	561
Total	213	426	639

Crash Threshold: -0.25%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	3	3	6
No Crash	101	532	633
Total	104	535	639

Crash Threshold: -0.50%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	0	0	0
No Crash	0	639	639
Total	0	639	639

Table H.9.1.7: Out-of-Sample Linear Crash Estimator LDA Confusion Matrices for Shell Stock (January 2008)

Crash Threshold: -0.10%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	267	151	418
No Crash	436	299	735
Total	703	450	1153

Crash Threshold: -0.25%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	127	40	167
No Crash	518	468	986
Total	645	508	1153

Crash Threshold: -0.50%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	33	11	44
No Crash	550	559	1109
Total	583	570	1153

Table H.9.1.8: Out-of-Sample Linear Crash Estimator LDA Confusion Matrices for Shell Stock (February 2008)

Crash Threshold: -0.10%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	149	120	269
No Crash	353	434	787
Total	502	554	1056

Crash Threshold: -0.25%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	32	12	44
No Crash	407	605	1012
Total	439	617	1056

Crash Threshold: -0.50%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	2	1	3
No Crash	318	735	1053
Total	320	736	1056

Table H.9.1.9: Out-of-Sample Linear Crash Estimator LDA Confusion Matrices for Shell Stock (March 2008)

Crash Threshold: -0.10%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	78	146	224
No Crash	195	453	648
Total	273	599	872

Crash Threshold: -0.25%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	14	24	38
No Crash	239	595	834
Total	253	619	872

Crash Threshold: -0.50%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	2	0	2
No Crash	137	733	870
Total	139	733	872

Table H.9.1.10: Out-of-Sample Linear Crash Estimator LDA Confusion Matrices for Shell Stock (April 2008)

Crash Threshold: -0.10%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	131	110	241
No Crash	384	506	890
Total	515	616	1131

Crash Threshold: -0.25%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	15	19	34
No Crash	390	707	1097
Total	405	726	1131

Crash Threshold: -0.50%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	0	3	3
No Crash	388	740	1128
Total	388	743	1131

Table H.9.1.11: Out-of-Sample Linear Crash Estimator LDA Confusion Matrices for Shell Stock (May 2008)

Crash Threshold: -0.10%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	114	58	172
No Crash	422	361	783
Total	536	419	955

Crash Threshold: -0.25%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	22	18	40
No Crash	388	527	915
Total	410	545	955

Crash Threshold: -0.50%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	1	7	8
No Crash	338	609	947
Total	339	616	955

Table H.9.1.12: Out-of-Sample Linear Crash Estimator LDA Confusion Matrices for Shell Stock (June 2008)

Crash Threshold: -0.10%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	164	150	314
No Crash	308	426	734
Total	472	576	1048

Crash Threshold: -0.25%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	55	49	104
No Crash	407	537	944
Total	462	586	1048

Crash Threshold: -0.50%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	5	4	9
No Crash	393	646	1039
Total	398	650	1048

Table H.9.2.1: Out-of-Sample VPIN LDA Confusion Matrices for Shell Stock (July 2007)

Crash Threshold: -0.10%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	91	94	185
No Crash	500	425	925
Total	591	519	1110

Crash Threshold: -0.25%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	32	15	47
No Crash	585	478	1063
Total	617	493	1110

Crash Threshold: -0.50%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	1	4	5
No Crash	328	777	1105
Total	329	781	1110

Table H.9.2.2: Out-of-Sample VPIN LDA Confusion Matrices for Shell Stock (August 2007)

Crash Threshold: -0.10%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	145	114	259
No Crash	450	431	881
Total	595	545	1140

Crash Threshold: -0.25%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	23	16	39
No Crash	536	565	1101
Total	559	581	1140

Crash Threshold: -0.50%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	2	4	6
No Crash	580	554	1134
Total	582	558	1140

Table H.9.2.3: Out-of-Sample VPIN LDA Confusion Matrices for Shell Stock (September 2007)

Crash Threshold: -0.10%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	67	46	113
No Crash	448	372	820
Total	515	418	933

Crash Threshold: -0.25%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	6	4	10
No Crash	482	441	923
Total	488	445	933

Crash Threshold: -0.50%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	0	0	0
No Crash	0	933	933
Total	0	933	933

Table H.9.2.4: Out-of-Sample VPIN LDA Confusion Matrices for Shell Stock (October 2007)

Crash Threshold: -0.10%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	114	59	173
No Crash	604	458	1062
Total	718	517	1235

Crash Threshold: -0.25%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	3	20	23
No Crash	485	727	1212
Total	488	747	1235

Crash Threshold: -0.50%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	0	0	0
No Crash	294	941	1235
Total	294	941	1235

Table H.9.2.5: Out-of-Sample VPIN LDA Confusion Matrices for Shell Stock (November 2007)

Crash Threshold: -0.10%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	136	171	307
No Crash	388	441	829
Total	524	612	1136

Crash Threshold: -0.25%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	35	34	69
No Crash	516	551	1067
Total	551	585	1136

Crash Threshold: -0.50%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	2	2	4
No Crash	479	653	1132
Total	481	655	1136

Table H.9.2.6: Out-of-Sample VPIN LDA Confusion Matrices for Shell Stock (December 2007)

Crash Threshold: -0.10%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	48	30	78
No Crash	239	322	561
Total	287	352	639

Crash Threshold: -0.25%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	3	3	6
No Crash	298	335	633
Total	301	338	639

Crash Threshold: -0.50%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	0	0	0
No Crash	0	639	639
Total	0	639	639

Table H.9.2.7: Out-of-Sample VPIN LDA Confusion Matrices for Shell Stock (January 2008)

Crash Threshold: -0.10%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	256	162	418
No Crash	442	293	735
Total	698	455	1153

Crash Threshold: -0.25%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	109	58	167
No Crash	573	413	986
Total	682	471	1153

Crash Threshold: -0.50%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	30	14	44
No Crash	557	552	1109
Total	587	566	1153

Table H.9.2.8: Out-of-Sample VPIN LDA Confusion Matrices for Shell Stock (February 2008)

Crash Threshold: -0.10%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	134	135	269
No Crash	395	392	787
Total	529	527	1056

Crash Threshold: -0.25%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	21	23	44
No Crash	537	475	1012
Total	558	498	1056

Crash Threshold: -0.50%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	1	2	3
No Crash	396	657	1053
Total	397	659	1056

Table H.9.2.9: Out-of-Sample VPIN LDA Confusion Matrices for Shell Stock (March 2008)

Crash Threshold: -0.10%

		Predicted		
		Crash	No Crash	Total
Actual	Crash	116	108	224
	No Crash	337	311	648
	Total	453	419	872

Crash Threshold: -0.25%

		Predicted		
		Crash	No Crash	Total
Actual	Crash	14	24	38
	No Crash	440	394	834
	Total	454	418	872

Crash Threshold: -0.50%

		Predicted		
		Crash	No Crash	Total
Actual	Crash	1	1	2
	No Crash	450	420	870
	Total	451	421	872

Table H.9.2.10: Out-of-Sample VPIN LDA Confusion Matrices for Shell Stock (April 2008)

Crash Threshold: -0.10%

		Predicted		
		Crash	No Crash	Total
Actual	Crash	134	107	241
	No Crash	494	396	890
	Total	628	503	1131

Crash Threshold: -0.25%

		Predicted		
		Crash	No Crash	Total
Actual	Crash	21	13	34
	No Crash	600	497	1097
	Total	621	510	1131

Crash Threshold: -0.50%

		Predicted		
		Crash	No Crash	Total
Actual	Crash	1	2	3
	No Crash	505	623	1128
	Total	506	625	1131

Table H.9.2.11: Out-of-Sample VPIN LDA Confusion Matrices for Shell Stock (May 2008)

Crash Threshold: -0.10%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	104	68	172
No Crash	455	328	783
Total	559	396	955

Crash Threshold: -0.25%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	18	22	40
No Crash	383	532	915
Total	401	554	955

Crash Threshold: -0.50%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	4	4	8
No Crash	331	616	947
Total	335	620	955

Table H.9.2.12: Out-of-Sample VPIN LDA Confusion Matrices for Shell Stock (June 2008)

Crash Threshold: -0.10%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	151	163	314
No Crash	395	339	734
Total	546	502	1048

Crash Threshold: -0.25%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	57	47	104
No Crash	515	429	944
Total	572	476	1048

Crash Threshold: -0.50%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	6	3	9
No Crash	507	532	1039
Total	513	535	1048

Table H.9.3.1: Out-of-Sample Market Heat LDA Confusion Matrices for Shell Stock (July 2007)

Crash Threshold: -0.10%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	111	74	185
No Crash	503	422	925
Total	614	496	1110

Crash Threshold: -0.25%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	33	14	47
No Crash	744	319	1063
Total	777	333	1110

Crash Threshold: -0.50%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	2	3	5
No Crash	492	613	1105
Total	494	616	1110

Table H.9.3.2: Out-of-Sample Market Heat LDA Confusion Matrices for Shell Stock (August 2007)

Crash Threshold: -0.10%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	150	109	259
No Crash	493	388	881
Total	643	497	1140

Crash Threshold: -0.25%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	22	17	39
No Crash	576	525	1101
Total	598	542	1140

Crash Threshold: -0.50%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	2	4	6
No Crash	385	749	1134
Total	387	753	1140

Table H.9.3.3: Out-of-Sample Market Heat LDA Confusion Matrices for Shell Stock (September 2007)

Crash Threshold: -0.10%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	60	53	113
No Crash	345	475	820
Total	405	528	933

Crash Threshold: -0.25%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	4	6	10
No Crash	439	484	923
Total	443	490	933

Crash Threshold: -0.50%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	0	0	0
No Crash	0	933	933
Total	0	933	933

Table H.9.3.4: Out-of-Sample Market Heat LDA Confusion Matrices for Shell Stock (October 2007)

Crash Threshold: -0.10%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	91	82	173
No Crash	484	578	1062
Total	575	660	1235

Crash Threshold: -0.25%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	10	13	23
No Crash	530	682	1212
Total	540	695	1235

Crash Threshold: -0.50%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	0	0	0
No Crash	317	918	1235
Total	317	918	1235

Table H.9.3.5: Out-of-Sample Market Heat LDA Confusion Matrices for Shell Stock (November 2007)

Crash Threshold: -0.10%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	129	178	307
No Crash	329	500	829
Total	458	678	1136

Crash Threshold: -0.25%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	32	37	69
No Crash	423	644	1067
Total	455	681	1136

Crash Threshold: -0.50%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	2	2	4
No Crash	346	786	1132
Total	348	788	1136

Table H.9.3.6: Out-of-Sample Market Heat LDA Confusion Matrices for Shell Stock (December 2007)

Crash Threshold: -0.10%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	29	49	78
No Crash	163	398	561
Total	192	447	639

Crash Threshold: -0.25%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	2	4	6
No Crash	322	311	633
Total	324	315	639

Crash Threshold: -0.50%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	0	0	0
No Crash	0	639	639
Total	0	639	639

Table H.9.3.7: Out-of-Sample Market Heat LDA Confusion Matrices for Shell Stock (January 2008)

Crash Threshold: -0.10%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	195	223	418
No Crash	339	396	735
Total	534	619	1153

Crash Threshold: -0.25%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	76	91	167
No Crash	408	578	986
Total	484	669	1153

Crash Threshold: -0.50%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	20	24	44
No Crash	393	716	1109
Total	413	740	1153

Table H.9.3.8: Out-of-Sample Market Heat LDA Confusion Matrices for Shell Stock (February 2008)

Crash Threshold: -0.10%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	125	144	269
No Crash	336	451	787
Total	461	595	1056

Crash Threshold: -0.25%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	20	24	44
No Crash	422	590	1012
Total	442	614	1056

Crash Threshold: -0.50%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	2	1	3
No Crash	468	585	1053
Total	470	586	1056

Table H.9.3.9: Out-of-Sample Market Heat LDA Confusion Matrices for Shell Stock (March 2008)

Crash Threshold: -0.10%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	66	158	224
No Crash	178	470	648
Total	244	628	872

Crash Threshold: -0.25%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	27	11	38
No Crash	471	363	834
Total	498	374	872

Crash Threshold: -0.50%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	0	2	2
No Crash	228	642	870
Total	228	644	872

Table H.9.3.10: Out-of-Sample Market Heat LDA Confusion Matrices for Shell Stock (April 2008)

Crash Threshold: -0.10%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	116	125	241
No Crash	438	452	890
Total	554	577	1131

Crash Threshold: -0.25%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	13	21	34
No Crash	555	542	1097
Total	568	563	1131

Crash Threshold: -0.50%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	3	0	3
No Crash	396	732	1128
Total	399	732	1131

Table H.9.3.11: Out-of-Sample Market Heat LDA Confusion Matrices for Shell Stock (May 2008)

Crash Threshold: -0.10%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	91	81	172
No Crash	370	413	783
Total	461	494	955

Crash Threshold: -0.25%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	24	16	40
No Crash	436	479	915
Total	460	495	955

Crash Threshold: -0.50%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	6	2	8
No Crash	274	673	947
Total	280	675	955

Table H.9.3.12: Out-of-Sample Market Heat LDA Confusion Matrices for Shell Stock (June 2008)

Crash Threshold: -0.10%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	130	184	314
No Crash	284	450	734
Total	414	634	1048

Crash Threshold: -0.25%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	46	58	104
No Crash	357	587	944
Total	403	645	1048

Crash Threshold: -0.50%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	4	5	9
No Crash	380	659	1039
Total	384	664	1048

Table H.10.1.1: Out-of-Sample Linear Crash Estimator LDA Confusion Matrices for Vodafone Stock (July 2007)

Crash Threshold: -0.10%

Actual		Predicted: Crash	Predicted: No Crash	Total
	Crash	163	96	259
	No Crash	493	392	885
	Total	656	488	1144

Crash Threshold: -0.25%

Actual		Predicted: Crash	Predicted: No Crash	Total
	Crash	46	18	64
	No Crash	680	400	1080
	Total	726	418	1144

Crash Threshold: -0.50%

Actual		Predicted: Crash	Predicted: No Crash	Total
	Crash	2	3	5
	No Crash	220	919	1139
	Total	222	922	1144

Table H.10.1.2: Out-of-Sample Linear Crash Estimator LDA Confusion Matrices for Vodafone Stock (August 2007)

Crash Threshold: -0.10%

Actual		Predicted: Crash	Predicted: No Crash	Total
	Crash	105	139	244
	No Crash	274	646	920
	Total	379	785	1164

Crash Threshold: -0.25%

Actual		Predicted: Crash	Predicted: No Crash	Total
	Crash	35	38	73
	No Crash	287	804	1091
	Total	322	842	1164

Crash Threshold: -0.50%

Actual		Predicted: Crash	Predicted: No Crash	Total
	Crash	5	4	9
	No Crash	228	927	1155
	Total	233	931	1164

Table H.10.1.3: Out-of-Sample Linear Crash Estimator LDA Confusion Matrices for Vodafone Stock (September 2007)

Crash Threshold: -0.10%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	95	97	192
No Crash	323	451	774
Total	418	548	966

Crash Threshold: -0.25%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	13	13	26
No Crash	391	549	940
Total	404	562	966

Crash Threshold: -0.50%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	0	2	2
No Crash	196	768	964
Total	196	770	966

Table H.10.1.4: Out-of-Sample Linear Crash Estimator LDA Confusion Matrices for Vodafone Stock (October 2007)

Crash Threshold: -0.10%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	152	109	261
No Crash	466	537	1003
Total	618	646	1264

Crash Threshold: -0.25%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	33	19	52
No Crash	400	812	1212
Total	433	831	1264

Crash Threshold: -0.50%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	3	0	3
No Crash	192	1069	1261
Total	195	1069	1264

Table H.10.1.5: Out-of-Sample Linear Crash Estimator LDA Confusion Matrices for Vodafone Stock (November 2007)

Crash Threshold: -0.10%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	153	158	311
No Crash	394	457	851
Total	547	615	1162

Crash Threshold: -0.25%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	34	37	71
No Crash	394	697	1091
Total	428	734	1162

Crash Threshold: -0.50%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	1	4	5
No Crash	209	948	1157
Total	210	952	1162

Table H.10.1.6: Out-of-Sample Linear Crash Estimator LDA Confusion Matrices for Vodafone Stock (December 2007)

Crash Threshold: -0.10%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	45	96	141
No Crash	134	393	527
Total	179	489	668

Crash Threshold: -0.25%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	7	16	23
No Crash	127	518	645
Total	134	534	668

Crash Threshold: -0.50%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	0	0	0
No Crash	0	668	668
Total	0	668	668

Table H.10.1.7: Out-of-Sample Linear Crash Estimator LDA Confusion Matrices for Vodafone Stock (January 2008)

Crash Threshold: -0.10%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	233	161	394
No Crash	428	345	773
Total	661	506	1167

Crash Threshold: -0.25%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	97	63	160
No Crash	438	569	1007
Total	535	632	1167

Crash Threshold: -0.50%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	28	18	46
No Crash	379	742	1121
Total	407	760	1167

Table H.10.1.8: Out-of-Sample Linear Crash Estimator LDA Confusion Matrices for Vodafone Stock (February 2008)

Crash Threshold: -0.10%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	138	130	268
No Crash	345	453	798
Total	483	583	1066

Crash Threshold: -0.25%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	26	26	52
No Crash	351	663	1014
Total	377	689	1066

Crash Threshold: -0.50%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	2	4	6
No Crash	193	867	1060
Total	195	871	1066

Table H.10.1.9: Out-of-Sample Linear Crash Estimator LDA Confusion Matrices for Vodafone Stock (March 2008)

Crash Threshold: -0.10%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	96	168	264
No Crash	190	426	616
Total	286	594	880

Crash Threshold: -0.25%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	41	56	97
No Crash	256	527	783
Total	297	583	880

Crash Threshold: -0.50%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	6	8	14
No Crash	219	647	866
Total	225	655	880

Table H.10.1.10: Out-of-Sample Linear Crash Estimator LDA Confusion Matrices for Vodafone Stock (April 2008)

Crash Threshold: -0.10%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	121	158	279
No Crash	372	504	876
Total	493	662	1155

Crash Threshold: -0.25%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	31	41	72
No Crash	418	665	1083
Total	449	706	1155

Crash Threshold: -0.50%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	3	2	5
No Crash	295	855	1150
Total	298	857	1155

Table H.10.1.11: Out-of-Sample Linear Crash Estimator LDA Confusion Matrices for Vodafone Stock (May 2008)

Crash Threshold: -0.10%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	75	143	218
No Crash	202	537	739
Total	277	680	957

Crash Threshold: -0.25%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	22	21	43
No Crash	544	370	914
Total	566	391	957

Crash Threshold: -0.50%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	1	2	3
No Crash	178	776	954
Total	179	778	957

Table H.10.1.12: Out-of-Sample Linear Crash Estimator LDA Confusion Matrices for Vodafone Stock (June 2008)

Crash Threshold: -0.10%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	148	136	284
No Crash	357	426	783
Total	505	562	1067

Crash Threshold: -0.25%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	40	38	78
No Crash	457	532	989
Total	497	570	1067

Crash Threshold: -0.50%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	3	7	10
No Crash	337	720	1057
Total	340	727	1067

Table H.10.2.1: Out-of-Sample VPIN LDA Confusion Matrices for Vodafone Stock (July 2007)

Crash Threshold: -0.10%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	152	107	259
No Crash	477	408	885
Total	629	515	1144

Crash Threshold: -0.25%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	30	34	64
No Crash	549	531	1080
Total	579	565	1144

Crash Threshold: -0.50%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	1	4	5
No Crash	563	576	1139
Total	564	580	1144

Table H.10.2.2: Out-of-Sample VPIN LDA Confusion Matrices for Vodafone Stock (August 2007)

Crash Threshold: -0.10%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	113	131	244
No Crash	454	466	920
Total	567	597	1164

Crash Threshold: -0.25%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	46	27	73
No Crash	533	558	1091
Total	579	585	1164

Crash Threshold: -0.50%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	6	3	9
No Crash	540	615	1155
Total	546	618	1164

Table H.10.2.3: Out-of-Sample VPIN LDA Confusion Matrices for Vodafone Stock (September 2007)

Crash Threshold: -0.10%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	112	80	192
No Crash	409	365	774
Total	521	445	966

Crash Threshold: -0.25%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	15	11	26
No Crash	489	451	940
Total	504	462	966

Crash Threshold: -0.50%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	0	2	2
No Crash	394	570	964
Total	394	572	966

Table H.10.2.4: Out-of-Sample VPIN LDA Confusion Matrices for Vodafone Stock (October 2007)

Crash Threshold: -0.10%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	159	102	261
No Crash	583	420	1003
Total	742	522	1264

Crash Threshold: -0.25%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	24	28	52
No Crash	614	598	1212
Total	638	626	1264

Crash Threshold: -0.50%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	1	2	3
No Crash	675	586	1261
Total	676	588	1264

Table H.10.2.5: Out-of-Sample VPIN LDA Confusion Matrices for Vodafone Stock (November 2007)

Crash Threshold: -0.10%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	159	152	311
No Crash	430	421	851
Total	589	573	1162

Crash Threshold: -0.25%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	35	36	71
No Crash	490	601	1091
Total	525	637	1162

Crash Threshold: -0.50%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	2	3	5
No Crash	602	555	1157
Total	604	558	1162

Table H.10.2.6: Out-of-Sample VPIN LDA Confusion Matrices for Vodafone Stock (December 2007)

Crash Threshold: -0.10%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	67	74	141
No Crash	242	285	527
Total	309	359	668

Crash Threshold: -0.25%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	10	13	23
No Crash	324	321	645
Total	334	334	668

Crash Threshold: -0.50%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	0	0	0
No Crash	303	365	668
Total	303	365	668

Table H.10.2.7: Out-of-Sample VPIN LDA Confusion Matrices for Vodafone Stock (January 2008)

Crash Threshold: -0.10%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	193	201	394
No Crash	370	403	773
Total	563	604	1167

Crash Threshold: -0.25%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	96	64	160
No Crash	578	429	1007
Total	674	493	1167

Crash Threshold: -0.50%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	26	20	46
No Crash	637	484	1121
Total	663	504	1167

Table H.10.2.8: Out-of-Sample VPIN LDA Confusion Matrices for Vodafone Stock (February 2008)

Crash Threshold: -0.10%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	145	123	268
No Crash	441	357	798
Total	586	480	1066

Crash Threshold: -0.25%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	32	20	52
No Crash	533	481	1014
Total	565	501	1066

Crash Threshold: -0.50%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	4	2	6
No Crash	530	530	1060
Total	534	532	1066

Table H.10.2.9: Out-of-Sample VPIN LDA Confusion Matrices for Vodafone Stock (March 2008)

Crash Threshold: -0.10%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	153	111	264
No Crash	319	297	616
Total	472	408	880

Crash Threshold: -0.25%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	57	40	97
No Crash	400	383	783
Total	457	423	880

Crash Threshold: -0.50%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	7	7	14
No Crash	438	428	866
Total	445	435	880

Table H.10.2.10: Out-of-Sample VPIN LDA Confusion Matrices for Vodafone Stock (April 2008)

Crash Threshold: -0.10%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	131	148	279
No Crash	413	463	876
Total	544	611	1155

Crash Threshold: -0.25%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	43	29	72
No Crash	555	528	1083
Total	598	557	1155

Crash Threshold: -0.50%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	0	5	5
No Crash	555	595	1150
Total	555	600	1155

Table H.10.2.11: Out-of-Sample VPIN LDA Confusion Matrices for Vodafone Stock (May 2008)

Crash Threshold: -0.10%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	140	78	218
No Crash	453	286	739
Total	593	364	957

Crash Threshold: -0.25%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	20	23	43
No Crash	314	600	914
Total	334	623	957

Crash Threshold: -0.50%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	1	2	3
No Crash	223	731	954
Total	224	733	957

Table H.10.2.12: Out-of-Sample VPIN LDA Confusion Matrices for Vodafone Stock (June 2008)

Crash Threshold: -0.10%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	171	113	284
No Crash	412	371	783
Total	583	484	1067

Crash Threshold: -0.25%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	43	35	78
No Crash	544	445	989
Total	587	480	1067

Crash Threshold: -0.50%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	5	5	10
No Crash	532	525	1057
Total	537	530	1067

Table H.10.3.1: Out-of-Sample Market Heat LDA Confusion Matrices for Vodafone Stock (July 2007)

Crash Threshold: -0.10%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	123	136	259
No Crash	380	505	885
Total	503	641	1144

Crash Threshold: -0.25%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	32	32	64
No Crash	688	392	1080
Total	720	424	1144

Crash Threshold: -0.50%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	2	3	5
No Crash	651	488	1139
Total	653	491	1144

Table H.10.3.2: Out-of-Sample Market Heat LDA Confusion Matrices for Vodafone Stock (August 2007)

Crash Threshold: -0.10%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	93	151	244
No Crash	331	589	920
Total	424	740	1164

Crash Threshold: -0.25%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	26	47	73
No Crash	380	711	1091
Total	406	758	1164

Crash Threshold: -0.50%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	4	5	9
No Crash	336	819	1155
Total	340	824	1164

Table H.10.3.3: Out-of-Sample Market Heat LDA Confusion Matrices for Vodafone Stock (September 2007)

Crash Threshold: -0.10%

Actual		**Predicted** Crash	No Crash	Total
	Crash	80	112	192
	No Crash	357	417	774
	Total	437	529	966

Crash Threshold: -0.25%

Actual		**Predicted** Crash	No Crash	Total
	Crash	12	14	26
	No Crash	330	610	940
	Total	342	624	966

Crash Threshold: -0.50%

Actual		**Predicted** Crash	No Crash	Total
	Crash	0	2	2
	No Crash	398	566	964
	Total	398	568	966

Table H.10.3.4: Out-of-Sample Market Heat LDA Confusion Matrices for Vodafone Stock (October 2007)

Crash Threshold: -0.10%

Actual		**Predicted** Crash	No Crash	Total
	Crash	114	147	261
	No Crash	396	607	1003
	Total	510	754	1264

Crash Threshold: -0.25%

Actual		**Predicted** Crash	No Crash	Total
	Crash	30	22	52
	No Crash	626	586	1212
	Total	656	608	1264

Crash Threshold: -0.50%

Actual		**Predicted** Crash	No Crash	Total
	Crash	0	3	3
	No Crash	579	682	1261
	Total	579	685	1264

Table H.10.3.5: Out-of-Sample Market Heat LDA Confusion Matrices for Vodafone Stock (November 2007)

Crash Threshold: -0.10%

Actual		Predicted: Crash	Predicted: No Crash	Total
	Crash	172	139	311
	No Crash	495	356	851
	Total	667	495	1162

Crash Threshold: -0.25%

Actual		Predicted: Crash	Predicted: No Crash	Total
	Crash	42	29	71
	No Crash	601	490	1091
	Total	643	519	1162

Crash Threshold: -0.50%

Actual		Predicted: Crash	Predicted: No Crash	Total
	Crash	2	3	5
	No Crash	553	604	1157
	Total	555	607	1162

Table H.10.3.6: Out-of-Sample Market Heat LDA Confusion Matrices for Vodafone Stock (December 2007)

Crash Threshold: -0.10%

Actual		Predicted: Crash	Predicted: No Crash	Total
	Crash	54	87	141
	No Crash	198	329	527
	Total	252	416	668

Crash Threshold: -0.25%

Actual		Predicted: Crash	Predicted: No Crash	Total
	Crash	8	15	23
	No Crash	165	480	645
	Total	173	495	668

Crash Threshold: -0.50%

Actual		Predicted: Crash	Predicted: No Crash	Total
	Crash	0	0	0
	No Crash	432	236	668
	Total	432	236	668

Table H.10.3.7: Out-of-Sample Market Heat LDA Confusion Matrices for Vodafone Stock (January 2008)

Crash Threshold: -0.10%

Actual		**Predicted** Crash	No Crash	Total
	Crash	172	222	394
	No Crash	318	455	773
	Total	490	677	1167

Crash Threshold: -0.25%

Actual		**Predicted** Crash	No Crash	Total
	Crash	72	88	160
	No Crash	410	597	1007
	Total	482	685	1167

Crash Threshold: -0.50%

Actual		**Predicted** Crash	No Crash	Total
	Crash	21	25	46
	No Crash	405	716	1121
	Total	426	741	1167

Table H.10.3.8: Out-of-Sample Market Heat LDA Confusion Matrices for Vodafone Stock (February 2008)

Crash Threshold: -0.10%

Actual		**Predicted** Crash	No Crash	Total
	Crash	119	149	268
	No Crash	311	487	798
	Total	430	636	1066

Crash Threshold: -0.25%

Actual		**Predicted** Crash	No Crash	Total
	Crash	22	30	52
	No Crash	377	637	1014
	Total	399	667	1066

Crash Threshold: -0.50%

Actual		**Predicted** Crash	No Crash	Total
	Crash	3	3	6
	No Crash	423	637	1060
	Total	426	640	1066

Table H.10.3.9: Out-of-Sample Market Heat LDA Confusion Matrices for Vodafone Stock (March 2008)

Crash Threshold: -0.10%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	87	177	264
No Crash	210	406	616
Total	297	583	880

Crash Threshold: -0.25%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	39	58	97
No Crash	322	461	783
Total	361	519	880

Crash Threshold: -0.50%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	7	7	14
No Crash	417	449	866
Total	424	456	880

Table H.10.3.10: Out-of-Sample Market Heat LDA Confusion Matrices for Vodafone Stock (April 2008)

Crash Threshold: -0.10%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	156	123	279
No Crash	448	428	876
Total	604	551	1155

Crash Threshold: -0.25%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	34	38	72
No Crash	526	557	1083
Total	560	595	1155

Crash Threshold: -0.50%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	3	2	5
No Crash	455	695	1150
Total	458	697	1155

Table H.10.3.11: Out-of-Sample Market Heat LDA Confusion Matrices for Vodafone Stock (May 2008)

Crash Threshold: -0.10%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	90	128	218
No Crash	265	474	739
Total	355	602	957

Crash Threshold: -0.25%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	26	17	43
No Crash	511	403	914
Total	537	420	957

Crash Threshold: -0.50%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	1	2	3
No Crash	412	542	954
Total	413	544	957

Table H.10.3.12: Out-of-Sample Market Heat LDA Confusion Matrices for Vodafone Stock (June 2008)

Crash Threshold: -0.10%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	163	121	284
No Crash	432	351	783
Total	595	472	1067

Crash Threshold: -0.25%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	48	30	78
No Crash	577	412	989
Total	625	442	1067

Crash Threshold: -0.50%

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	4	6	10
No Crash	575	482	1057
Total	579	488	1067

Appendix I

Table I.1: Multiple Linear Regression of Lagged Variables on Out-of-Sample FX Returns

(Monthly Sampling)

	United Kingdom			Eurozone			Switzerland		
variable	**coefficient**	**p-value**	**R^2**	**coefficient**	**p-value**	**R^2**	**coefficient**	**p-value**	**R^2**
constant	0.0164	0.0054	0.0424	-			0.0159	0.0188	-
Lagged Returns	-	-					-	-	
Interest Rate Premium	-	-					-	-	
Inflation	-	-					-	-	
Unemployment	-	-					-	-	
Current Account	-	-					-	-	
Reserves	0.1741	0.0081					-	-	
Money Supply (M2)	-	-					-	-	
GDP Growth	-	-					-	-	
Index Returns	-	-					-	-	
TED-Spread	-3.7865 e-4	0.0067					-3.2792 e-4	0.0404	
Δ VIX	-	-					-	-	

Table I.1 (Continued): Multiple Linear Regression of Lagged Variables on Out-of-Sample FX Returns

(Monthly Sampling)

	Australia			Canada			Japan		
variable	**coefficient**	**p-value**	**R^2**	**coefficient**	**p-value**	**R^2**	**coefficient**	**p-value**	**R^2**
constant	-	-		-	-		-	-	
Lagged Returns	-	-		-	-		-	-	
Interest Rate Premium	-	-		-	-		-	-	
Inflation	-	-		-	-		-	-	
Unemployment	-	-		-	-		-	-	
Current Account	-4.0236	0.0000	0.0857	-	-	0	-	-	0
Reserves	-	-		-	-		0.1314	0.0000	
Money Supply (M2)	-	-		-	-		-	-	
GDP Growth	-	-		-	-		-	-	
Index Returns	-	-		-	-		-	-	
TED-Spread	-	-		-	-		-	-	
Δ VIX	-	-		-0.0243	0.0000		-	-	

Table I.1 (Continued): Multiple Linear Regression of Lagged Variables on Out-of-Sample FX Returns

(Monthly Sampling)

	Sweden			Norway			New Zealand		
variable	**coefficient**	**p-value**	**R^2**	**coefficient**	**p-value**	**R^2**	**coefficient**	**p-value**	**R^2**
constant	0.0024	0.0027	0	0.0168	0.0153	0	0.0212	0.0063	0
Lagged Returns	-	-		-	-		-	-	
Interest Rate Premium	-	-		-	-		-	-	
Inflation	-	-		-	-		-	-	
Unemployment	-	-		-	-		-	-	
Current Account	-	-		-	-		-	-	
Reserves	-	-		-	-		-	-	
Money Supply (M2)	-0.4065	9.5635 e-4		-	-		-	-	
GDP Growth	-	-		-	-		-	-	
Index Returns	-	-		-	-		-	-	
TED-Spread	-4.7275 e-4	0.0033		-3.5196 e-4	0.0310		-4.9540 e-4	0.0070	
Δ VIX	-	-		-	-		-	-	

Appendix J

Table J.1: Signals Approach Out-of-Sample Confusion Matrices for United Kingdom

1 or More Signals

Actual		Predicted: Crash	Predicted: No Crash	Total
	Crash	13	4	17
	No Crash	40	20	60
	Total	53	24	77

2 or More Signals

Actual		Predicted: Crash	Predicted: No Crash	Total
	Crash	10	7	17
	No Crash	20	40	60
	Total	30	47	77

3 or More Signals

Actual		Predicted: Crash	Predicted: No Crash	Total
	Crash	10	7	17
	No Crash	11	49	60
	Total	21	56	77

4 or More Signals

Actual		Predicted: Crash	Predicted: No Crash	Total
	Crash	7	10	17
	No Crash	6	54	60
	Total	13	64	77

Table J.2: Signals Approach Out-of-Sample Confusion Matrices for Eurozone

1 or More Signals

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	14	3	17
No Crash	46	14	60
Total	60	17	77

2 or More Signals

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	11	6	17
No Crash	31	29	60
Total	42	35	77

3 or More Signals

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	9	8	17
No Crash	20	40	60
Total	29	48	77

4 or More Signals

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	6	11	17
No Crash	11	49	60
Total	17	60	77

Table J.3: Signals Approach Out-of-Sample Confusion Matrices for Switzerland

1 or More Signals

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	18	1	19
No Crash	50	8	58
Total	68	9	77

2 or More Signals

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	12	7	19
No Crash	29	29	58
Total	41	36	77

3 or More Signals

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	10	9	19
No Crash	19	39	58
Total	29	48	77

4 or More Signals

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	5	14	19
No Crash	8	50	58
Total	13	64	77

Table J.4: Signals Approach Out-of-Sample Confusion Matrices for Australia

1 or More Signals

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	13	1	14
No Crash	57	6	63
Total	70	7	77

2 or More Signals

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	10	4	14
No Crash	40	23	63
Total	50	27	77

3 or More Signals

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	10	4	14
No Crash	26	37	63
Total	36	41	77

4 or More Signals

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	4	10	14
No Crash	11	52	63
Total	15	62	77

Table J.5: Signals Approach Out-of-Sample Confusion Matrices for Canada

1 or More Signals

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	10	4	14
No Crash	45	18	63
Total	55	22	77

2 or More Signals

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	8	6	14
No Crash	20	43	63
Total	28	49	77

3 or More Signals

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	7	7	14
No Crash	8	55	63
Total	15	62	77

4 or More Signals

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	4	10	14
No Crash	3	60	63
Total	7	70	77

Table J.6: Signals Approach Out-of-Sample Confusion Matrices for Japan

1 or More Signals

Actual		Predicted: Crash	Predicted: No Crash	Total
	Crash	14	0	14
	No Crash	40	23	63
	Total	54	23	77

2 or More Signals

Actual		Predicted: Crash	Predicted: No Crash	Total
	Crash	8	6	14
	No Crash	25	38	63
	Total	33	44	77

3 or More Signals

Actual		Predicted: Crash	Predicted: No Crash	Total
	Crash	7	7	14
	No Crash	12	51	63
	Total	19	58	77

4 or More Signals

Actual		Predicted: Crash	Predicted: No Crash	Total
	Crash	7	7	14
	No Crash	10	53	63
	Total	17	60	77

Table J.7: Signals Approach Out-of-Sample Confusion Matrices for Sweden

1 or More Signals

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	14	4	18
No Crash	41	18	59
Total	55	22	77

2 or More Signals

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	8	10	18
No Crash	27	32	59
Total	35	42	77

3 or More Signals

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	5	13	18
No Crash	13	46	59
Total	18	59	77

4 or More Signals

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	4	14	18
No Crash	8	51	59
Total	12	65	77

Table J.8: Signals Approach Out-of-Sample Confusion Matrices for Norway

1 or More Signals

Actual		Predicted: Crash	Predicted: No Crash	Total
	Crash	19	2	21
	No Crash	44	12	56
	Total	63	14	77

2 or More Signals

Actual		Predicted: Crash	Predicted: No Crash	Total
	Crash	14	7	21
	No Crash	34	22	56
	Total	48	29	77

3 or More Signals

Actual		Predicted: Crash	Predicted: No Crash	Total
	Crash	8	13	21
	No Crash	15	41	56
	Total	23	54	77

4 or More Signals

Actual		Predicted: Crash	Predicted: No Crash	Total
	Crash	8	13	21
	No Crash	6	50	56
	Total	14	63	77

Table J.9: Signals Approach Out-of-Sample Confusion Matrices for New Zealand

1 or More Signals

Actual		Predicted: Crash	Predicted: No Crash	Total
	Crash	13	5	18
	No Crash	51	8	59
	Total	64	13	77

2 or More Signals

Actual		Predicted: Crash	Predicted: No Crash	Total
	Crash	11	7	18
	No Crash	35	24	59
	Total	46	31	77

3 or More Signals

Actual		Predicted: Crash	Predicted: No Crash	Total
	Crash	8	10	18
	No Crash	16	43	59
	Total	24	53	77

4 or More Signals

Actual		Predicted: Crash	Predicted: No Crash	Total
	Crash	7	11	18
	No Crash	9	50	59
	Total	16	61	77

Appendix K

Table K.1: Binary In-Sample Crash Estimation Results (January 2000 – June 2006)[30]

country	variable	logit	variable	probit
United Kingdom	constant	-1.859406 (0.3687632)	constant	-1.750648 (0.3397966)
	VIX	3.438467 (1.753634)	TED-Spread	0.0168237 (0.0072853)
			Reserves	-8.802038 (4.006543)
Eurozone	constant	-1.570555 (0.3417863)	constant	-0.9211882 (0.1825515)
	VIX	6.397665 (2.1587)	VIX	3.682519 (1.21522)
	Index Return	17.31092 (7.22152)	Index Return	9.872468 (4.064429)
Switzerland	constant	-1.187166 (0.2710327)	constant	-0.7264997 (0.1584402)
Australia	constant	-2.230034 (0.5615126)	constant	-1.363707 (0.3308798)
	TED-Spread	0.029967 (0.0123843)	TED-Spread	0.0183853 (0.0074885)
Canada	constant	-2.47092 (0.4279363)	constant	-1.419188 (0.2109854)
Japan	constant	-0.9162907 (0.2522625)	constant	-0.5659488 (0.1524533)
Sweden	constant	-2.892443 (0.682691)	constant	-1.709441 (0.3734137)
	TED-Spread	0.0372642 (0.0131957)	TED-Spread	0.0220537 (0.0076819)
	Money Supply	31.55654 (12.69668)	Money Supply	18.65052 (7.492772)
	Reserves	-19.76751 (8.555822)	Reserves	-11.67296 (4.831465)
Norway	constant	-1.393379 (0.2882487)	constant	-0.8483243 (0.164433)
	VIX	3.406396 (1.402876)	VIX	2.107986 (0.8344955)
New Zealand	constant	-1.187166 (0.2710327)	constant	-0.7264997 (0.1584402)

[30] All values in parentheses throughout Appendix H show respective p-values for each variable.

Appendix L

Table L.1: Pooled Logit In-Sample Estimation (January 2000 – June 2006)

Logistic Regression

Number of obs.	= 693
Wald chi2 (2)	= 30.30
Prob. > chi2	= 0.0000
Pseudo R2	= 0.0383

Log pseudolikelihood = -344.39245

crash	Coefficient	Robust Standard Errors	z	P > |z|	[95% Confidence Interval]	
TED-Spread	.0184287	.0042494	4.34	0.000	.0101001	.0267573
VIX	1.480911	.4785649	3.09	0.002	.5429415	2.418881
constant	-2.069384	.1938379	-10.68	0.000	-2.449299	-1.689468

Table L.2: Pooled Probit In-Sample Estimation (January 2000 – June 2006)

Probit Regression

Number of obs.	= 693
Wald chi2 (2)	= 30.53
Prob. > chi2	= 0.0000
Pseudo R2	= 0.0396

Log pseudolikelihood = -343.92422

crash	Coefficient	Robust Standard Error	z	P > |z|	[95% Confidence Interval]	
TED-Spread	.0111228	.0025391	4.38	0.000	.0061462	.0160995
VIX	.8922877	.28439	3.14	0.002	.3348935	1.449682
constant	-1.251406	.1122989	-11.14	0.000	-1.471508	-1.031304

Appendix M

Table M.1: Country-by-Country Out-of-Sample Logit Confusion Matrices

United Kingdom

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	3	14	17
No Crash	3	58	61
Total	6	72	78

Eurozone

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	3	14	17
No Crash	6	55	61
Total	9	69	78

Switzerland

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	0	19	19
No Crash	0	59	59
Total	0	78	78

Australia

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	6	8	14
No Crash	16	48	64
Total	22	56	78

Canada

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	0	14	14
No Crash	0	64	64
Total	0	78	78

Japan

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	0	15	15
No Crash	0	63	63
Total	0	78	78

Sweden

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	6	12	18
No Crash	20	40	60
Total	26	52	78

Norway

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	4	17	21
No Crash	4	53	57
Total	8	70	78

New Zealand

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	0	18	18
No Crash	0	60	60
Total	0	78	78

Appendix N

Table N.1: Country-by-Country Out-of-Sample Probit Confusion Matrices

United Kingdom

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	6	11	17
No Crash	6	55	61
Total	12	72	78

Eurozone

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	3	14	17
No Crash	6	55	61
Total	9	69	78

Switzerland

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	0	19	19
No Crash	0	59	59
Total	0	78	78

Australia

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	6	8	14
No Crash	16	48	64
Total	22	56	78

Canada

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	0	14	14
No Crash	0	64	64
Total	0	78	78

Japan

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	0	15	15
No Crash	0	63	63
Total	0	78	78

Sweden

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	6	12	18
No Crash	20	40	60
Total	26	52	78

Norway

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	4	17	21
No Crash	4	53	57
Total	8	70	78

New Zealand

Actual	Predicted: Crash	Predicted: No Crash	Total
Crash	0	18	18
No Crash	0	60	60
Total	0	78	78

Appendix O

Table O.1: Pooled Logit Out-of-Sample Confusion Matrix

		Predicted		
		Crash	No Crash	Total
Actual	Crash	45	108	153
	No Crash	54	495	549
	Total	99	603	702

Table O.2: Pooled Probit Out-of-Sample Confusion Matrix

		Predicted		
		Crash	No Crash	Total
Actual	Crash	45	108	153
	No Crash	54	495	549
	Total	99	603	702

Appendix P

Table P.1: Random Effects Model (January 2000 – June 2006)

Random-effects GLS Regression			Number of obs.		= 693
Group variable: **country**			Number of groups		= 9
R^2:	within	= 0.1022	Obs. per group:	min	= 77
	between	= 0.3037		avg	= 77.0
	overall	= 0.1025		max	= 77
correlation (ε_i, X)		= 0 (assumed)	Wald chi²(**5**)		= 9251.16
(correlation between explanatory variables and error terms)			Prob > chi²		= 0.0000

(Standard Errors adjusted for 9 clusters in country)

crash	Coefficient	Robust Standard Error	z	P > |z|	[95% Confidence Interval]	
Δ interest rate premium	-.0107598	.0029774	-3.61	0.000	-.0165954	-.0049243
Δ current account/GDP	-.0030491	.0001748	-17.44	0.000	-.0033917	-.0027064
GDP growth	.2122208	.0653019	3.25	0.001	.0842315	.3402101
VIX	-.0245184	.0038531	-6.36	0.000	-.0320703	-.0169665
TED-Spread	-.0003332	.0000374	-8.90	0.000	-.0004065	-.0002598
constant	.0143798	.001421	10.12	0.000	.0115948	.0171648
σ_u	0	(standard deviation of residuals within countries)				
σ_e	.02698606	(standard deviation of residuals)				
ρ	0	(fraction of variance due to differences across countries)				

Table P.2: Random Effects Model – Unit Root Test

Levin-Lin-Chu unit-root test for **fxreturn**

Ho: Panels contain unit roots		Number of panels =	**9**
Ha: Panels are stationary		Number of periods =	**77**
AR parameter:	**Common**	Asymptotics: **N/T -> 0**	
Panel means:	**Included**		
Time trend:	**Not included**		
ADF regressions:	**0.67** lags average (chosen by AIC)		
LR variance:	**Bartlett** kernel, **13.00** lags average (chosen by **LLC**)		

	Statistic	p-value
Unadjusted t	**-22.3161**	
Adjusted t*	**-20.7402**	**0.0000**

Table P.3: Fixed Effects Model (January 2000 – June 2006)

Fixed-effects (within) Regression			Number of obs.		= 693
Group variable: **country**			Number of groups		= 9
R^2:	within	= 0.1022	Obs. per group:	min	= 77
	between	= 0.2974		avg	= 77.0
	overall	= 0.1025		max	= 77
correlation (ε_i, X)		= 0.0119	Wald chi^2(**5**)		= 3696.46
(correlation between explanatory variables and error terms)			Prob. > chi^2		= 0.0000

(Std. Err. adjusted for 9 clusters in country)

crash	Coefficient	Robust Standard Errors	z	P > |z|	[95% Confidence Interval]	
Δ interest rate premium	-.0107147	.0029884	-3.59	0.007	-.0176058	-.0038235
Δ current account/GDP	-.0031157	.0001708	-18.24	0.000	-.0035096	-.0027218
GDP growth	.2028469	.0743218	2.73	0.026	.0314605	.3742334
VIX	-.0245315	.0038306	-6.40	0.000	-.0333648	-.0156981
TED-Spread	-.0003329	.0000377	-8.84	0.000	-.0004198	-.0002461
constant	.0144088	.0013422	10.74	0.000	.0113138	.0175039
σ_μ	.00120878	(standard deviation of residuals within groups)				
σ_e	.02698606	(standard deviation of residuals)				
ρ	.00200239	(fraction of variance due to differences across countries)				

Table P.4: Fixed Effects Model – Unit Root Test

Levin-Lin-Chu unit-root test for **fxreturn**

Ho: Panels contain unit roots		Number of panels =	**9**
Ha: Panels are stationary		Number of periods =	**77**
AR parameter:	**Common**	Asymptotics: **N/T -> 0**	
Panel means:	**Included**		
Time trend:	**Not included**		
ADF regressions:	**0.67** lags average (chosen by AIC)		
LR variance:	**Bartlett** kernel, **13.00** lags average (chosen by **LLC**)		

	Statistic	p-value
Unadjusted t	**-22.3161**	
Adjusted t*	**-20.7402**	**0.0000**