Macroscopic Look at Equity Markets

by

Abdullah AlShelahi

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Industrial and Operations Engineering)
in The University of Michigan
2019

Doctoral Committee:

Professor Romesh Saigal, Chair
Associate Professor Eunshin Byon
Assistant Professor Cong Shi
Professor Ji Zhu

Abdullah AlShelahi

shelahi@umich.edu

ORCID iD: 0000-0002-9774-0771

Dedication

This dissertation is dedicated to my parents and everyone who has believed in me and supported me along this journey.

Acknowledgments

I would like to thank Prof. Romesh Saigal for chairing my dissertation committee and for his insightful comments and suggestions. It was my pleasure to know him as a person and as an advisor. Thanks again for sharing your infinite wisdom in academia and life. This dissertation wouldn't have been a success without your mentoring and help throughout the past several years.

I would like to especially thank Prof. Ji Zhu, Prof. Eunshin Byon, and Prof. Cong Shi for agreeing to be on my dissertation committee. They were generous with their expertise and time, and provided helpful and excellent feedback. It was my privilege to know them. During the last two years, I have had the opportunity to work closely with Prof. Byon on two projects in which I truly benefited from her insights and knowledge. I would like to express my warmest thanks to Prof. Philip Roe and Prof. Karthik Duraisamy from the department of Aerospace Engineering for teaching me concepts and ideas in Computational Fluid Dynamics. Special thanks go to the staff and faculty members of the department of Industrial and Operations Engineering for their continued encouragement and support during my graduate years.

Many thanks to all my friends in Michigan who have made my graduate years wonderful. My heartfelt thanks to Prof. Christine Feak, Elizabeth and Tarey for helping me, tremendously, to be a better writer and storyteller. Also, I'm especially grateful for the scholarship provided by Kuwait University.

To my family, thank you for the sacrifices and understanding during the last few years. Mom and Dad, you have been very supportive and patient at every step in my education. Thus, I dedicate this dissertation to you.

Table of Contents

List of Figures

Figure

List of Tables

Table

Abstract

Financial markets contribute to the stability of the global economy. A vivid example of this crucial connection is the crash in 2008. Although this connection is well-established, the underlying structure of markets is complex. Complex systems such as this tend to operate in a nonlinear fashion, generating extreme and rare events. In the era of high-frequency trading, we have been witnessing other unusual and extreme phenomena such as flash crashes and technical difficulties due to glitches. Given the presence of these phenomena, a powerful tool is needed to monitor markets' behavior and activities. Recent research in finance has mainly focused on analyzing individual stocks (i.e., microscopic analysis) while ignoring the overall interactions and dynamics between them (i.e., macroscopic analysis), which is crucial to identify abnormalities. To better represent markets' behavior, this dissertation proposes a novel macroscopic perspective, allowing for prediction and understanding of anomalies. Although, the focus of this dissertation is mainly on equity (or stocks) markets, the proposed methods are general such that it can be extended to other financial markets.

In the first part of this dissertation, we propose new sensors to monitor equity markets, adopting a macroscopic perspective. This perspective offers new insights into the physics of stocks. Briefly, we analyze stocks within the context of fluid dynamics in which the movement of stocks is viewed as particles within the "fluid flow" context. For the first time, concepts from physics are utilized and incorporated in modeling both internal and external dynamics. Based on the physics of fluid dynamics, we develop macroscopic variables, such as density, velocity, flux, and pressure. A model consisting of a system of stochastic nonlinear partial differential equations is introduced. The model connects and determines the evolution of macroscopic variables. The validation and usefulness of the new approach is discussed in the

last part of this chapter.

The second part extends the analysis to examine the structural properties of the proposed model. We show that the model exhibits weak solutions, such as shock and rarefaction waves. These solutions provide a new narrative about financial shocks. We also present a theoretical analysis of the behavior of the macroscopic variables. To solve the system of stochastic nonlinear partial differential equations adaptively, we devise an integrative algorithm which combines numerical methods and stochastic filtering techniques. This algorithm is tested on abnormal and normal trading days. The results suggest that abnormalities can be identified.

In the third part, we tackle the problem of detecting medium intensity crashes, which are macroscopic abnormalities occurring on a given trading day. These abnormalities create an imbalance with normal market activities because they rarely occur. To address these challenges, we present a cost-sensitive classification model based on a recurrent neural-network called Reservoir Computing (RC). We also extend the classical RC to process predictions from the physics-based model presented in Chapter 2. This approach includes information about the underlying mechanism of the markets, showing significant improvement in detection accuracy.

This dissertation offers a macroscopic perspective by incorporating internal and external dynamics. Market makers can use this perspective to detect irregularities and questionable practices. When market makers are more aware, regular investors will no longer worry that they are trading in a losing or rigged game.

Chapter 1

Introduction

Financial markets, including equity [1] and derivatives markets, are experiencing serious challenges that are often characterized by a sharp decline during crisis periods. In the last 20 years, we have witnessed several financial crises with devastating consequences. The crisis in 2008 is a testimony to the complexity of the financial system. This complexity is increasing, exponentially, with the advent and expansion of computing technology and high-frequency trading [61]. Though the benefits of such technologies are profound, abnormal phenomena have been observed, including the so-called flash crash on May 6th, 2010 [112] and technical difficulties (or glitches), such as the 2012 Facebook shares listing incident. A flash crash is defined as a sudden drop that lasts for a short period of time [52]. These crashes may evolve in the future [38]. Another implication of the adaption of new trading practices is the creation of unjust opportunities [84]. Thus, monitoring markets is necessary to detect and predict abnormal activities while ensuring fairness and equal opportunities in trading. Decision makers should adopt a perspective that appreciates the existing challenges that are attributable to the complex interactions between internal and external influences and forces. However, currently, decision makers lack an effective monitoring tool.

Despite the development of models in finance and economics, the behavior of financial markets is still an open question that is considered an exciting example of a system's com-

[1] In this dissertation, the terms "stocks" and "equity" are used interchangeably.

plexity [65]. The authors of [65] assert that several impractical assumptions about market dynamics have persisted in the literature. For instance, although the equilibrium theory is the baseline of many economics models, financial markets are not in equilibrium. As many other critics assert, markets remain unbalanced which means that the general assumptions of fair play can be called into question [73]. Another example of questionable assumptions can be seen in the finance literature. The Efficient Market Hypothesis (EMH) is a well-known hypothesis in which rational traders are competing to maximize their profits in a market where asset prices reflect all the available information. A wide range of papers criticize the EMH and reject the assumption that all traders are completely rational [75, 85, 46]. In fact, financial markets, where several participants are competing to gain the most profit, is considered a progressive system. These markets, generally, exhibit irrationality and unusual patterns.

Looking more broadly at the prediction literature, we can see two main streams emerge. One assumes crises are *Black Swans* and thus can not be detected [94]. The other main stream suggests that markets exhibit signals prior to a crisis (or *dragon-king* and tipping point), [111], drawing heavily from physics literature. Given the powerful insights that can emerge from physics, this dissertation is related to the latter main stream, aiming to propose models that accurately represent markets' behavior and possibly predict anomalies [2].

If we look at the most successful methods, concepts, and theories in finance or economics to date, we may be surprised to observe the substantial quantity of significant works that have been derived (partially or wholly) from or solved by the aid of physics. In a complex and stochastic system where smaller system entities interact, it is essential that the conventional ways of thinking be avoided, as they may lead to irrational actions. Physics can come into play here by suggesting laws for tackling complex systems through an appreciation of the interaction of these systems' entities. Our approach in this dissertation is to use physics, in particular fluid dynamics, to model the stochastic and macroscopic nature of equity markets.

[2]The terms "anomaly" and "abnormality " are used interchangeably throughout this dissertation.

Though we focus our analysis on equity markets, our approach can be extended to other markets. Extensions and future works are highlighted in Chapter 5.

In equity markets, we can think of each stock as an entity in a larger, more complex system, with many other entities. A plethora of research in this area concentrates on the study and analysis of individual stock movements, and the movements of a particular portfolio of those individual stocks [25, 122, 129, 107]. Inspired by fluid dynamics, we study the flow of stocks in the markets by defining mass as the number of stocks at a specific time and price, and density as the mass per unit price. We build a dynamical model to understand the flow of all stocks among all markets, not only a single market. Unlike other models, our model is macroscopic and constitutes a stochastic version of conservation laws from physics. These laws are composed of a system of nonlinear partial differential equations. In this dissertation, we investigate a stock as a contributor to the mass in the price-time space. We mimic studies found in oceanography, where for example, the application of conservation principles explains the creation of ocean waves. This contrasts with a microscopic look at the ocean, where the movement of each water particle is considered [8].

This dissertation introduces an original perspective about equity markets. A macroscopic model from the viewpoint of fluid dynamics and physics is proposed. This model can alert regulators to unusual activities in the flow of stocks. In particular, the fluid flow in physics is used to measure the irregularities. The established link between physics and finance will bring greater clarity to those who run markets, as well as regular investors who will no longer worry that they are trading in a losing or rigged game.

1.1 Summary of Contributions

This dissertation is presented in three separate scientific manuscripts. In Chapter 2 and 3, we incorporate material from [1] and [2], both coauthored with Romesh Saigal. The contribution and organization of this dissertation is summarized below.

Chapter 2 builds the foundation for the adaption of fluid dynamics concepts to equity

markets. We create new measures for the detection and prediction of unusual events, based on ideas from the physics of fluid dynamics. We then rigorously derive and interpret the mass and momentum conservation laws as they apply to equity markets. A new model consisting of a system of nonlinear stochastic differential equations is proposed. Three main properties of the proposed model are highlighted. The model is fitted to flash crash data and the results are analyzed.

In Chapter 3, we investigate further the properties of the proposed model. The type of solutions generated by the model are shown to be in the form of shock and rarefaction waves. We link these solutions to financial shocks and provide conditions under which they are formed. To solve the proposed model dynamically, we propose an efficient algorithm, which is a novel integration of numerical methods with stochastic filtering techniques. The results of implementing the proposed algorithm on two datasets reveal that unusual market flows can be identified.

Chapter 4 studies the prediction and detection of medium intensity crashes (i.e, market crashes in the trading day). Studies dealing with these crashes are scarce in the literature. By considering medium intensity crashes as the minority class of market activities, we propose a cost-sensitive classification approach based on a neural network structure, known as Reservoir Computing (RC). RC is used to classify markets as normal or abnormal. We further introduce a hybrid classification approach which combines the classical RC and the physics-based model introduced in Chapter 2 and further investigated in Chapter 3. To test the hybrid approach, we utilize unique datasets from the previous ten years. The proposed hybrid approach demonstrates that it can detect abnormalities with high Precision and Recall.

The remainder of this dissertation presents the details of our research as outlined above. Each chapter is presented in a scientific format, with its own introduction, methods, results, and conclusion sections. The dissertation concludes with chapter 5, which provides a summary of this research and highlights future directions.

Chapter 2

Insights into the Macroscopic Behavior of Equity Markets: Theory and Applications

2.1 Introduction

Equity market crashes are defined as sharp declines in the value of market securities. The crashes are rare and hard to explain. In 1987, the US stock market dropped by about thirty percent over the course of four trading days. We may wonder what causes financial crashes and whether they signal market inefficiency. The research in this area is still growing, but fundamental aspects of crashes are poorly understood, including how to predict them [74][1].

As market exchanges have become more automated, with the adoption of limit-order systems, and as the capability of computing technology expands, high-frequency trading will become a relatively common practice. With the advancement of methods of execution, new abnormalities and irregularities have started to appear. These events raise questions about the stability and structure of financial markets [72]. In the last decade, stock markets and

[1]The results of this chapter have been published in [2].

economies have experienced time periods that are characterized by sudden state transition and high unpredictability [112]

The flash crash is a poignant example of such irregularities. A flash crash is a mini-crash that is sudden and causes a sharp drop in the market within a short period of time [52]. The authors of [112] believe that flash crashes are not new phenomena, and that they in fact share similarities with previous crashes but exhibit different specifics. Given the aim of realizing profits by traders, one can expect more such crashes in the future [38]. An example of this is the recent flash crash of Amazon stock [17].

The manuscript [27] argues that a flash crash is a failure in a large-scale complex socio-technical system. In 2007, the Large-Scale Complex Systems IT Systems Initiative was established to address issues in science and technology related to financial-market failures. The initiative constitutes the interactions of the following fields: complexity science, predictable software systems, high-integrity systems engineering, socio-technical systems engineering, organizational complexity, and novel computation approaches. This program, which resulted in some publications on system failures, was terminated in 2013, but the problem is still relevant. Although many papers have investigated the reasons for mini-crashes, only a few have provided a suitable model for predicting them [52, 38].

The complex-systems approach, therefore, might provide insights for policy makers, academics, and market participants since the dynamic nature of complex systems is well suited to capture the nonlinear relationships between their inputs and outputs [112]. According to power-law statistics, small and large events belong to the same distribution, and it's believed that major catastrophes started out as small events that didn't stop growing and became extreme events [111]. The scientific community considers extreme events to be unpredictable and in many cases hard to forecast. This view is emphasized by Bak in [6], where the concept of self-organized criticality was introduced. The notion blossomed when Taleb introduced the black swan concept in 2007 [94]. However, a recent study of financial crashes [111] suggests that markets undergo a phase transition before experiencing a crash. This hypothesis also

suggests that equity markets exhibit signals referred to as dragon kings, which are extreme events that belong to the same class of certain other events but with an amplifying mechanism. In their article [125], Werner and his colleagues proposed examples where dragon kings can be detected. A recent method for detecting bubbles and crashes are reported in [123] where the authors show that a potential force is detected before a main crash in which it is considered as a "precursor" to a potential risk in the markets.

In this chapter, we investigate the existence of market sensors, which monitor certain market measures and send out impulses in cases of abnormal activity. These measures are based on properties drawn from physics and modeled as a dynamical system to determine their future evolution. Physics and finance have interacted for a long time. The one-dimensional random walk was introduced in finance by Bachelier in 1900 after the venture of physicists into stock markets [4]. Bachelier's work was eventually forgotten, but it was reinvestigated by scientists in 1950 and later to describe various aspects of financial markets. [119]. Their works have motivated significant research in options pricing theory, portfolio theory, and many other areas [119, 92]. In 1900 Bachelier introduced the notion of Brownian motion in his Ph.D. dissertation, and Einstein used probabilistic models to explain the theory [39]. Brownian motion was first discovered by the biologist Robert Brown in 1827 [15] while studying microscopic pollen particles floating in water. Geometric Brownian motion has become the central mathematical model in pricing financial derivatives [39]. In the finance literature, a stock price $S(t)$ is assumed to be a stochastic process and follows a geometric Brownian motion given by

$$dS(t) = \alpha S(t)dt + \sigma S(t)dW(t), \tag{2.1}$$

where α is the drift term, σ is the volatility of $S(t)$, and $W(t)$ is the standard Wiener process. The properties of a Wiener process can be found in any standard textbook on option pricing and financial mathematics.

In equation (2.1), we see an attempt to analyze a single random variable $S(t)$ in a dynamical system. This model is microscopic; it aims to model an individual stock and

study its behavior. Even if this model represents the actual behavior of the underlying stock price $S(t)$, it does not capture the overall dynamics of the stock market. The purpose of our study is to define a macroscopic model with sensors that are triggered by abnormal market activity.

The increased interest in financial modeling has motivated a new line of research, in mesoscopic models [56], which in many cases consider a portfolio of stocks. Market indexes are a good example of these models. Other mesoscopic models in statistical physics are used in modeling heterogeneous agents with a view to understanding their interactions. A well-known example of a mesoscopic model is a Fokker–Plank equation [56]. Several other methods and tools from statistical mechanics have been employed in financial markets [14, 21, 35]. They were implemented in the first place because they are capable of dealing with sophisticated environments where complex behavior arises from a rather simple interaction of some components. The purpose of this trend of research is to identify the universal and non-universal features of financial data [28]. A major area of research in the physics of finance is concerned with power tails, especially in wealth distribution, which was predicted by Pareto. The tools from statistical mechanics help to identify the features of tails in wealth distribution [22, 83]. There have also been attempts to derive models with the tools of the kinetic theory of fluids [29, 36]. Such models are based on partial differential equations (PDEs), which makes it possible to obtain general information and derive asymptotic behavior analytically [28]. An important application of fluid dynamics in the financial markets is introduced in [128]. There, the authors propose a new model for financial price movements in which the prices in the order book are described as colloidal Brownian particles.

Therefore, it is clear that the implementation of statistical mechanics and physics in finance is not new. Each method has its novelty and importance to the field. In both microscopic and mesoscopic models, the dynamical model, if it exists and is specified, is concerned with only a single stock or a portfolio of stocks. This approach might be useful if the aim is to study the stocks within the portfolio under study. An obvious drawback of such

models is that even though the PDEs involved might provide a good prediction of the studied state behavior, it might be hard to verify with data. Also, those models might not capture the overall behavior of markets, as they consider a small subset of that market. Further, research based on PDEs with agents as the central argument are lacking in information, assuming that such information is not available to the public.

In this chapter, we advance the literature by building a system of stochastic partial differential equations to analyze and monitor equity markets macroscopically. This model employs physics principles borrowed from fluid dynamics, and provides new measures for detection and prediction of unusual events. It begins by aggregating all stocks traded in several exchanges in an economic zone. It then defines their movement in the price–time domain through the definition of macroscopic variables such as density, average velocity, flux and pressure, the model then connects these variables through "conservation principals" to define a dynamical system represented by a stochastic partial differential equation (SPDE), and the sensors then monitor several of the resulting parameters defining the SPDE.

It is important to mention that the analogy between fluid dynamics and economics was investigated extensively in the literature (see e.g., [76, 88, 35, 29, 101]). This analogy was approached from different angles so that new results might emerge. The introduction of the hydraulic model in macroeconomics is a good example of such attempts. Recent research in this direction is reported in [31] where the authors propose a hydraulic model to explain a simple trade flow between two countries. Our model is adaptive to the conditions of the market, suggested by the data. This is another important distinguishing feature of our work.

Our model will benefit market makers and regulators, as it is a predictive model that can be used to monitoring equity markets. We envision that the model can be integrated with an automation system composed of an artificial intelligence algorithm that sends out an alarm in cases of abnormal behavior. One might ask: What is the best action to take when an alarm is received? This is a separate research question, which we will pursue later. In this work, we focus on building the dynamical model. The solution of our proposed model

is challenging, and will be covered in Chapter 3. The integration of the artificial intelligence which is capable of warning of the impending instability in the market is under investigation.

The chapter proceeds as follows. In section 2, we present the macroscopic model and review its properties. In section 3, we examine the assumptions made in our analysis, verify them with data, and present computational results. Section 4 outlines our future work in this area and concludes the chapter.

2.2 The Equity Markets Macroscopic Model

Microscopic models are common because they tend to be sensitive to small perturbations, but they fail to provide a quantitative description of macroscopic phenomena. The research problem we consider here is: Can we build sensors that alert market makers to abnormal activity? We investigate the existence of such sensors within a proposed novel macroscopic dynamical model by defining new market measures which can be sensed and thus give a different prospective on markets [1].

Observing a stock market from a monitor, one can see the movement of stock prices as fluid. We view the movement of the stocks in the market as the flow of fluid particles. In physics the flow of fluid particles is well understood, but considering a stock flow in this way needs further explanation. We can think of a stock as a molecule or particle that's flowing in a market exchange. The flow takes place in the price–time domain, unlike the three-dimensional space and time in physics. Each stock represents a single particle. Generally speaking, in physics there are two ways to describe the resulting fluid flow: the Eulerian description and the Lagrangian description. In the Lagrangian description, each fluid particle is tracked while it is moving in time and space, while the Eulerian description focuses on a specific sub-space and observes an aggregate particles of fluid as they flow through sub-space as time moves forward. In analogy of equity markets to fluid flow, in this work we focus on the Eulerian description.

The first macroscopic variable of the model is the density ρ, whose measure in number of stocks per unit price. For example, in traffic flow models, $\rho(x,t)$ is the density of cars on a highway, that is, the number of cars per kilometer at milepost x at time t [82]. In our model, $\rho(x,t)$ is the number of stocks per price x at time t.

Let $x \in \mathbb{R}^+$ represent the price of stocks, and let $\rho(x,t)$ be the density of stocks at price x and time t. The domain of space x and time t is discretized into regions $[x_1, x_2] \times [t_1, t_2]$, where $[x_1, x_2]$ is a price range (also called a price section) and $[t_1, t_2]$ is a time interval (See section 3.2 for the details.) Let $N(x,t)$ be the total number of stocks in section $[x_1, x_2]$ at time t. It is given by the integral of the density over $[x_1, x_2]$:

$$N(x,t) = \int_{x_1}^{x_2} \rho(x,t)dx \tag{2.2}$$

Another variable, the velocity $v_k(t)$ of stock k at time t, is defined as

$$v_k(t) = \lim_{\Delta t \to 0} \frac{p_k(t) - p_k(t - \Delta t)}{\Delta t}, \tag{2.3}$$

where $p_k(t)$ is the price of stock k at time t. The average velocity of stocks in price section $[x_1, x_2]$ at time t is denoted by $v(x,t)$, which is constant for each $x \in [x_1, x_2]$ and is defined as the average change in stock price in that price section at time t. The average velocity function is given by

$$v(x,t) = \frac{1}{N(x,t)} \sum_{k=1}^{N(x,t)} v_k(t). \tag{2.4}$$

If we assume that during the period of study (say a day), stocks are neither created nor destroyed, in the sense that initial public offerings (IPOs) and bankruptcies are rare and usually occur when the markets are usually closed, then the change in the number of stocks in price section $[x_1, x_2]$ is due only to the flow of stocks across the endpoints x_1, x_2. The stocks within that section are assumed to be compressible. Compressibility is a concept similar to its counterpart in fluid dynamics: The density of stocks (the number of stocks in

the domain parametrized by x and t) can change in response to external or internal forces. This assumption states that the number stocks given a price section might change as time proceeds.

As we assumed that stocks are neither created nor destroyed, the mass of stocks (number of stocks to be more precise) can change only as a result of flow of stocks across the endpoints x_1, x_2. The rate of flow (also called the flux) of stocks at point (x, t), Q(x,t), is given by

$$Q(x,t) = \rho(x,t)v(x,t) \tag{2.5}$$

By assuming that the boundary of the domain is closed, so that no stocks leave it or enter from outside, the rate of change in the number of stocks in $[x_1, x_2]$ is given by only the difference in flux at the endpoints x_1, x_2

$$\frac{d}{dt}\int_{x_1}^{x_2} \rho(x,t)dx = \rho(x_1,t)v(x_1,t) - \rho(x_2,t)v(x_2,t) \tag{2.6}$$

This equation is known as the law of mass conservation, and is a consequence the fact that no stocks are created or destroyed within the exchanges, but only move between the price–time regions. This is the integral form of the law of mass conservation. Another form is obtained by integrating equation (2.6) from t_1 to t_2. This form gives an expression for the number of stocks at time t_2 in terms of the fluxes at the endpoints x_1, x_2 and the mass of stocks at time t_1.

$$\int_{x_1}^{x_2} \rho(x,t_2)dx = \int_{x_1}^{x_2} \rho(x,t_1)dx + \int_{t_1}^{t_2} \rho(x_1,t)v(x_1,t)dt - \int_{t_1}^{t_2} \rho(x_2,t)v(x_2,t)dt \tag{2.7}$$

Assuming that the density $\rho(x,t)$ and average velocity $v(x,t)$ are differentiable functions, we can obtain the differential form of the law of conservation of mass:

$$\int_{t_1}^{t_2}\int_{x_1}^{x_2} \Big(\frac{\partial}{\partial t}\rho(x,t) + \frac{\partial}{\partial x}\Big(\rho(x,t)v(x,t)\Big)\Big)dxdt = 0 \tag{2.8}$$

Since (2.8) must hold for any price section $[x_1, x_2]$ and over any time interval $[t_1, t_2]$, the integrand of (8) must be equal to zero:

$$\frac{\partial}{\partial t}\rho(x,t) + \frac{\partial}{\partial x}\Big(\rho(x,t)v(x,t)\Big) = 0 \tag{2.9}$$

To get (2.9), we used the following:

$$\rho(x_2,t)v(t,x_2) - \rho(x_1,t)v(t,x_1) = \int_{x_1}^{x_2} \frac{\partial}{\partial x}\Big(\rho(x,t)v(t,x)\Big)dx \tag{2.10}$$

and

$$\rho(x,t_2) - \rho(x,t_1) = \int_{t_1}^{t_2} \frac{\partial}{\partial t}\rho(x,t)dt \tag{2.11}$$

Equation (2.9) expresses the conservation of mass and is often called the continuity equation. This equation gives the evolution of the density from the initial conditions. The partial differential equation (2.9) is nonlinear in the conserved quantity ρ, and has many solutions in the two global variables, ρ and v. In applications to transportation systems, there is a relationship between v and ρ. When the density of cars is high, their velocity is low. Equation (9) is solvable with this relationship [126, 127, 103]. In financial markets, there is no relationship between the number of stocks at a given price and their average velocity. Figure 2.1 shows scatter plots of density vs. velocity at four different times during the day of a flash crash. The plots show no evidence of a clear correlation between the two variables. The correlation coefficient is provided with each scatter plot. We see that the correlation coefficient is very low in three of the four times.

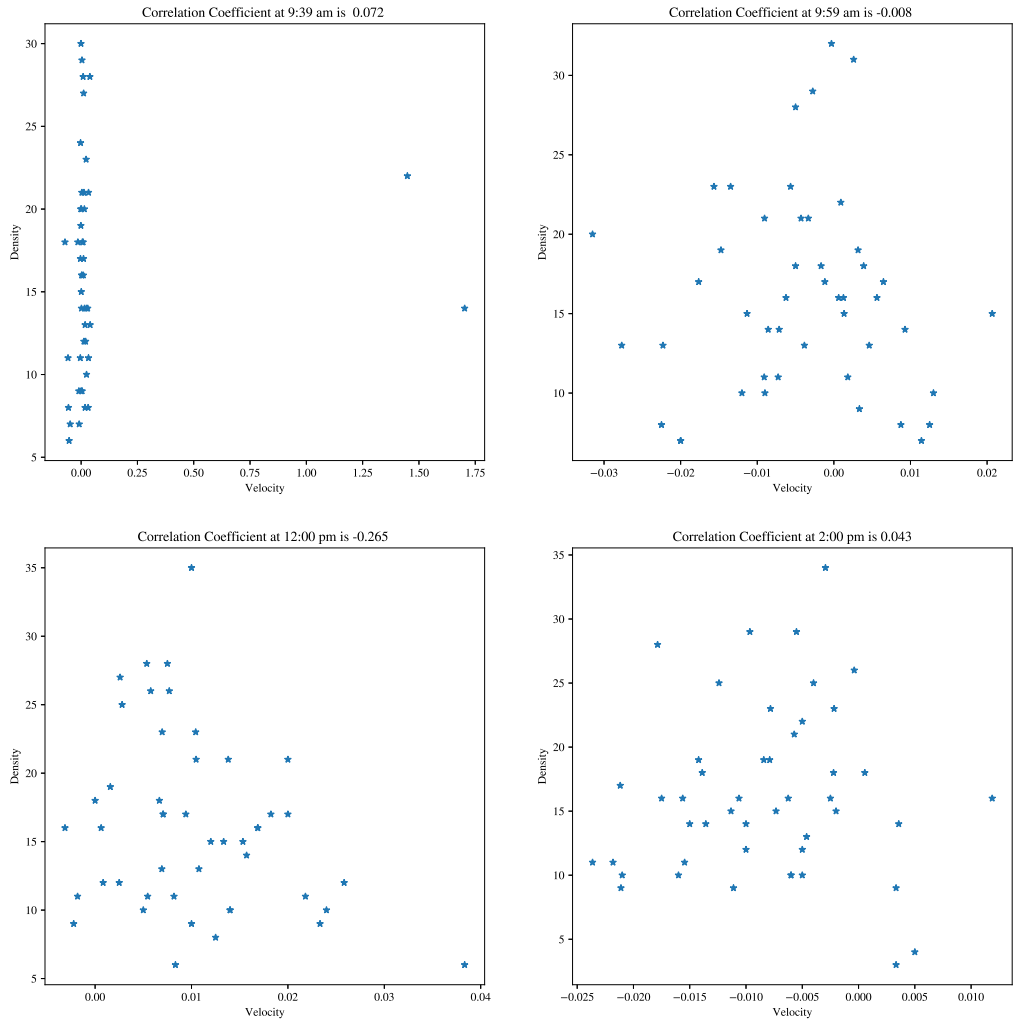


Figure 2.1: Scatter plots of density vs. velocity on the day of a flash crash

Another partial differential equation is necessary for the system of equations to have locally unique solutions. A good candidate is the so-called a momentum equation in fluid dynamics. Stocks don't collide and can cross each other, so momentum conservation of fluid dynamics in financial markets may not be exactly as in fluid flow. However, we assume that

a version of the momentum of stocks flow captures the market forces.

In this chapter, we define density as the number of stocks per unit price in contrast to the physical definition as the number of particles per unit space. The physical meaning of density is well understood.

To understand the financial meaning of density, let us consider the density at (x,t), $\rho(x,t)$, per unit price x and time t. Then [2]

$$\begin{aligned} x\rho(x,t) &= \text{number of stocks (per unit price) at price } x \times \text{ price per stock} \\ &= (\$/(\text{unit price})) \text{ in stocks priced } x \text{ at time } t \\ &= m(x,t) \text{ say.} \end{aligned}$$

Let the "flux" be $Q(x,t) = \Big(\rho(x,t)v(x,t)\Big)$, and consider:

$$\begin{aligned} \frac{\partial}{\partial t}m(x,t) + \frac{\partial}{\partial x}\Big(m(x,t)v(x,t)\Big) &= \frac{dx}{dt}\rho(x,t) + x\frac{\partial}{\partial t}\rho(x,t) + x\frac{\partial}{\partial x}Q(x,t) + Q(x,t) \\ &= Q(x,t) + x(\frac{\partial}{\partial t}\rho(x,t) + \frac{\partial}{\partial x}Q(x,t)) + Q(x,t) \\ &= 2Q(x,t). \end{aligned}$$

The above equation follows from the mass conservation which implies

$$\frac{\partial}{\partial t}\rho(x,t) + \frac{\partial}{\partial x}Q(x,t) = 0. \tag{2.12}$$

[2] $\int_{x_1}^{x_2} x\rho(x,t)dx = \$$ invested in stocks within $[x_1, x_2]$ at time t.

Let us consider the unit of the "flux" of stocks $Q(x,t)$:

$$Q(x,t) = \rho(x,t)v(x,t) = \frac{\text{Number of stocks } (\#)}{\text{unit price}} \ \frac{\text{Average change in stock prices } (\$)}{\text{unit time}}$$
$$= \frac{\text{Change in money}}{\text{unit price unit time}}$$

From the above, we can say that

$$Q(x,t)\,dxdt = \text{Change in money in } [x, x+dx] \times [t, t+dt]$$

or

$$\int_{t_1}^{t_2}\int_{x_1}^{x_2} Q(x,t)dxdt = \text{Change in money in } [x_1, x_2] \times [t_1, t_2] \tag{2.13}$$

In the equity markets, this change has two sources, one is the change due to market forces affecting all stocks, the other is external forces specific to the stock, which is generated by the flow of money in/out of the stock. The sum of these two forces acting on stocks are responsible for price movements. Overall, the change in money in price–time domain, even due to internal forces is hard to argue is conserved, but clearly the money is withdrawn/invested due to external force destroys this conservation. This can be formally expressed as follows:

$$\frac{\partial}{\partial t}\int_{x_1}^{x_2} Q(x,t)\,dx + \int_{x_1}^{x_2} \frac{\partial}{\partial x}\Big(Q(x,t)v(x,t)\Big)dx \neq 0 \tag{2.14}$$

Equation (2.14) follows from the nonconservation of the change in money, represented by the "flux", is not conserved. Therefore, an extra term is needed to account for these forces

acting on stocks,

$$\frac{\partial}{\partial t}\int_{x_1}^{x_2} Q(x,t)\,dx + \int_{x_1}^{x_2}\frac{\partial}{\partial x}\Big(Q(x,t)v(x,t)\Big)dx = -\int_{x_1}^{x_2}\frac{\partial}{\partial x}\theta(x,t)dx \tag{2.15}$$

where $\theta(x,t)$ [3] is the external force per unit price, say "Pressure", in price x at time t.

Integrate (15) from $t_1 \to t_2$,

$$\begin{aligned}\int_{t_1}^{t_2}\int_{x_1}^{x_2} Q(x,t)\,dxdt + \int_{x_1}^{x_2}\int_{t_1}^{t_2}\frac{\partial}{\partial x}\Big(Q(x,t)v(x,t)\Big)dxdt \\ = -\int_{x_1}^{x_2}\int_{t_1}^{t_2}\frac{\partial}{\partial x}\theta(x,t)dxdt\end{aligned} \tag{2.16}$$

or

$$\int_{t_1}^{t_2}\int_{x_1}^{x_2}\Big[\frac{\partial}{\partial t}Q(x,t) + \frac{\partial}{\partial x}\Big(Q(x,t)v(x,t)\Big) + \frac{\partial}{\partial x}\theta(x,t)\Big]dxdt = 0. \tag{2.17}$$

Equation (2.17) holds for any time and price intervals, we can conclude that the integrand in (17) must be equal to zero,

$$\frac{\partial}{\partial t}Q(x,t) + \frac{\partial}{\partial x}\Big(Q(x,t)v(x,t) + \theta(x,t)\Big) = 0 \tag{2.18}$$

From the above derivations, we see that the precise definition of momentum is not needed. However, for the sake of clarity, we will refer to Equation (12.18) as the conservation of momentum.

Equations (2.9) and (2.18) together constitute a system of partial differential equations:

$$\begin{aligned}\frac{\partial}{\partial t}\rho(x,t) + \frac{\partial}{\partial x}\Big(\rho(x,t)v(x,t)\Big) = 0 \\ \frac{\partial}{\partial t}Q(x,t) + \frac{\partial}{\partial x}\Big(Q(x,t)v(x,t) + \text{Pressure}\Big) = 0\end{aligned} \tag{2.19}$$

[3] This variable is hereafter referred to as "Pressure".

In addition to the system of equations in (19), an initial condition must be specified:

$$q(x,0) = q_0(x), \tag{2.20}$$

where $q \in \mathbb{R}^2$ is the vector of the conserved quantities, $q = (\rho, Q)^T$. The problem in equations (2.19) and (2.20) is called an initial-value problem or a Cauchy problem. Cauchy problems are often solved by the method of characteristic. Details about solving nonlinear partial differential equations can be found in [41, 82].

The behavior and shape of the pressure in equation (2.19) is analyzed through data. Our experiments on different days show that the pressure is a function of flux and velocity:

$$\text{Pressure}(x,t) = \alpha Q(x,t)v(x,t), \tag{2.21}$$

for some $\alpha \in \mathbb{R}$.

Systems of partial differential equations of the form (2.19) provide an excellent description of many models in physics, biology, and the social sciences. A PDE is said to be nonlinear if some term in the equation is nonlinear in at least one of the dependent variables and its partial derivatives (taken together). Nonlinear partial differential equations govern a broad range of complex systems. First-order systems of PDEs are those in which the order of highest-orderd of the derivative is one; they can be classified into three groups. These classes of equations model various phenomena and behavior. The system in (2.19), together with the pressure equation (2.21), makes that system of PDEs hyperbolic; a proof of this is given below. Hyperbolic PDEs are well suited to the financial markets, the knowledge and use of systems of nonlinear PDEs, like the one here, are growing. Also the numerical methods to solve them are well developed and readily implemented.

Theorem 2.1. *The system of partial differential equations in (2.19) is hyperbolic.*

Proof. Setting $q := (\rho, Q)^T$ and $f(q) = (q_2, (1+\alpha)\frac{q_2^2}{q_1})^T$, we can rewrite equation (2.19),

together with the pressure equation (2.21), in the form

$$\partial_t q + f(q)_x = 0,$$

or

$$\partial_t q + A(q)\partial_x q = 0, \tag{2.22}$$

where

$$A(q) = \partial f(q) := \begin{pmatrix} 0 & 1 \\ -(1+\alpha)\frac{q_2^2}{q_1^2} & 2(1+\alpha)\frac{q_2}{q_1} \end{pmatrix}$$

The eigenvalues of $A(q)$ are $\lambda_1 = (1+\alpha+\sqrt{\alpha^2+\alpha})\frac{q_2}{q_1}$ and $\lambda_2 = (1+\alpha-\sqrt{\alpha^2+\alpha})\frac{q_2}{q_1}$, which are real and distinct when $\alpha > 0$ or $\alpha < -1$, and $\frac{q_2}{q_1} \neq 0$.

□

The solution of a system of nonlinear partial differential equations involves wave propagation [82]. The system of nonlinear hyperbolic partial differential equations can admit discontinuities under smooth initial conditions. The discontinuities can take various forms, depending on the structure of the system's characteristics. The characteristics are the eigenvalues of $A(q)$.

Theorem 2.2. *The system of partial differential equations in (2.19) admits only shocks or rarefaction waves.*

We say that a solution ρ is a generalized solution if it satisfies (2.8) for all t_1, t_2, x_1, x_2. Equation (2.22) might not have a unique solution, and this happens when a shock forms. As a result, there is no classical solution to the system of PDEs, and one defines a weak solution.

Theorem 2.3. *The system of partial differential equations in (2.19) has a weak solution.*

Weak solutions are often not unique [80]. Thus an additional problem is to identify which solution is physical. The entropy condition states that a solution is physically relevant if the entropy increases across the shock. The properties of hyperbolic partial differential equations can be found in [41, 82].

2.2.1 The Stochastic Macroscopic Model

The PDEs can be written in the conservation form under the assumption that mass and momentum are conserved. The right-hand sides of both equations in (2.19) are zero, which asserts that fact. In equity markets, external forces might exist. IPOs and bankruptcy may occur rarely; however, a forcing term is necessary to account for the discretization error and sources of uncertainty that disturb the conservation of the flow. When dealing with data, there are sources of error that are inherent to the data collection, giving additional justification for the forcing terms of the form we have suggested. We propose adding drift and stochastic terms for that purpose. The modified model is as follows:

$$\frac{\partial}{\partial t}\rho(x,t) + \frac{\partial}{\partial x}\rho(x,t)v(x,t) = z_1(x,t) \tag{2.23}$$

$$\frac{\partial}{\partial t}\Big(\rho(x,t)v(x,t)\Big) + \frac{\partial}{\partial x}\Big(Q(x,t)v(x,t) + \alpha Q(x,t)v(x,t)\Big) = z_2(x,t) \tag{2.24}$$

$$z_1(x,t) = a(x,t) + b(x,t)\rho(x,t) + \sigma_1(x,t)\frac{dW_1(x,t)}{dxdt} \tag{2.25}$$

$$z_2(x,t) = c(x,t) + d(x,t)Q(x,t) + \sigma_2(x,t)\frac{dW_2(x,t)}{dxdt} \tag{2.26}$$

The right-hand sides of equations (2.25) and (2.26) contain deterministic functions $a(x,t)$, $b(x,t)$, $c(x,t)$, and $d(x,t)$ and Brownian sheets W. The functions $a(x,t)$ and $c(x,t)$ are included to capture the effect of the mean inflow and outflow [26]. The functions $\sigma_1(x,t)$ and $\sigma_2(x,t)$ represent the volatility of the process and capture the amount of disturbance to the conservation laws. The right-hand sides of equations (2.25) and (2.26) comprise reversion processes that revert to the mean value at rates of $b(x,t)$ and $d(x,t)$, respectively,

thus converging to the long-run means (as $t \to \infty$) [26].

A Brownian sheet $W(x,t)$ is a Gaussian stochastic process indexed by x and t with mean equal to 0 and covariance $\mathbf{E}\Big(W(x_1,t_1)W(x_2,t_2)\Big) = \min(x_1,x_2)\cdot\min(t_1,t_2)$. The random perpetuation $\frac{dW(x,t)}{dxdt}$ is white noise at (x,t) that helps in modeling the stochastic disturbance in the macroscopic model [126, 120]. We can interpret $dW(x,t)$ as

$$dW(x,t) = W(x+dx,t+dt) - W(x,t+dt) - W(x+dx,t) + W(x,t) \tag{2.27}$$

Given the properties of a Brownian sheet and the interpretation in (31), $dW(x,t)$ is a Gaussian stochastic process with mean zero and $\text{Var}[dW(x,t)] = dx\cdot dt$ [120].

Equations (2.23) through (2.26) are a system of stochastic partial differential equations. These equations determine the evolution of the macroscopic variables when uncertainty exists. The macroscopic variables are designed to capture the behavior of the flow of stocks, including possibly abnormal flows. Partial differential equations of this type admit shock waves, and we solve systems of this type numerically since analytical solutions to such systems are still being developed.

2.3 Numerical Results and Examples

The aim of this section is to verify the SPDE introduced in the previous section and, with the aid of data from a unique trading day, to analyze the macroscopic variables and verify the proposed SPDE. We derive the values of the macroscopic variables, density and velocity, from the data and investigate the values of the parameters on the right-hand sides of (2.25) and (2.26) that fit the model. From these results, we investigate which of these parameters may be good sensors for detecting unusual activity.

2.3.1 Data

The data used in this research are aggregated by-minute stock prices from three market exchanges (NASDAQ, AMEX, NYSE) or a total of 1000 stocks. A stock was included in the data only if it was actively traded at least 90 percent of the time. When data for a particular minute were not available, we made the assumption that the stock price at that time was the previously traded price. We analyzed data from the three markets on the day of a flash crash, May 6, 2010. On that day, the equity markets started to crash at 14:32, and the crash lasted about 36 minutes. The S&P Futures declined by 5% between 14:30 and 14:45. Various explanations for the flash crash were given [38], including the following:

- Technical difficulties at NYSE and ARCA
- Changes in the U.S. dollar/Japanese yen exchange rate
- Large purchase of put options
- Large sell of E-Mini contracts
- Quote stuffing (HFT)

Recently a trader from London has pleaded guilty in a US court to spoofing the US financial markets and is awaiting sentencing [12]. To get a mesoscopic look at the financial markets on that day, Figure 2.2 shows the relative performance of five financial indexes. The indexes exhibit similar behavior around the time of the flash crash; they each experience a sudden plunge followed by some instability. The flash crash phenomenon is unique, and that is the reason we chose that day. The incident caused tremendous losses within a span of few minutes. Many had believed that such an incident was impossible. However, we believe that this crash started earlier that day and could have been caught sooner.

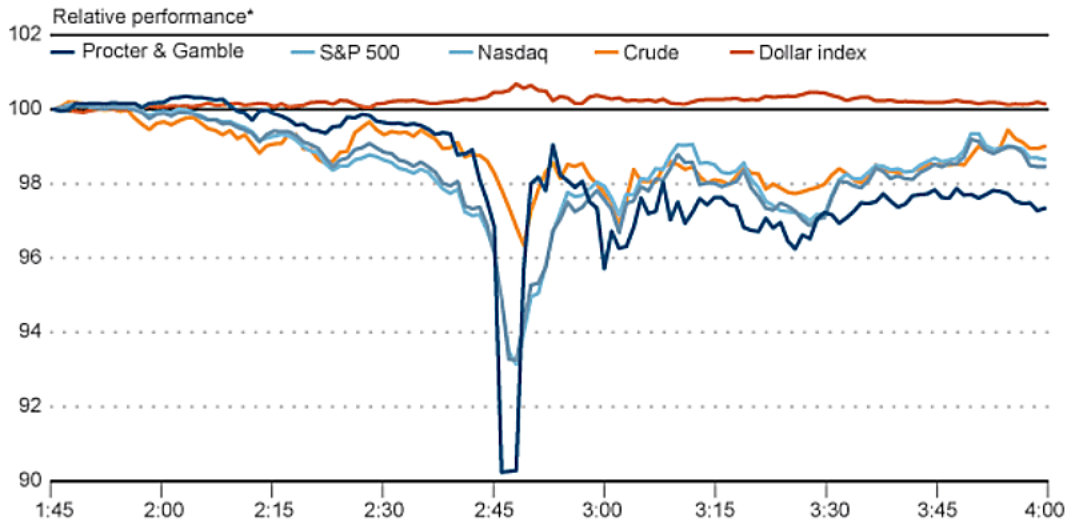


Figure 2.2: Flash Crash Chart [Thomson Reuters]

2.3.2 Domain Discretization

The macroscopic variables such as density $\rho(x,t)$ and velocity $v(x,t)$ were parametrized by price x and time t. The temporal domain t was divided into M intervals with a subinterval length $\Delta t = 1$ to match the data frequency (Figure 2.3). The stock market starts operating at 9:30 am and closes at 4:00 pm. Thus, the data correspond to a time period of 391 minutes, where $t_0 = 0$ and $t_M = 390$.

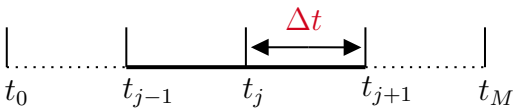


Figure 2.3: Time Discretization

The spatial domain, which is the stock price x, is divided into N intervals with a subinterval length Δx (Figure 2.4).

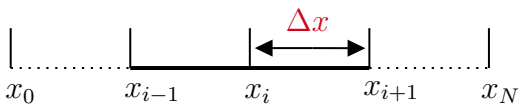


Figure 2.4: Space Discretization

In Figure 2.4, x_0 is set to a very small number, and x_N is set to an arbitrarily large number, for instance \$150, as few stocks are traded above this value. Thus most stock prices fall within the price range $[x_0, x_N]$. We define cell (i, j) as the small region $[x_i, x_{i+1}] \times [t_j, t_{j+1}]$ in the price–time domain, as shown in Figure 2.5. This domain construction is necessary to compute the macroscopic variables from the raw stock data. The length of a subinterval Δx can be either fixed or increasing. A good spatial domain discretization candidate is $\Delta x = \beta_1 x$ for some $\beta_1 \in \mathbb{R}^+$. The intuition behind this choice is that we expect to have fewer stocks as the price increases, so that having an increasing spatial domain with $\beta_1 \in [0, 1]$ would capture the highest number of stocks within a relatively small number of cells (Figure 2.6). A simpler discretization is $\Delta x = \beta_2$ for some $\beta_2 \in \mathbb{R}^+$. To get a sense of these discretization schemes, we computed the density of stocks for the day of the flash crash with $\beta_1 = 0.1$ and $\beta_2 = 1$, and plotted their heat maps. The density in cell $[x_i, x_{i+1}] \times [t_j, t_{j+1}]$ was defined as the number of stocks within the price section $[x_i, x_{i+1}]$ at time t_j.

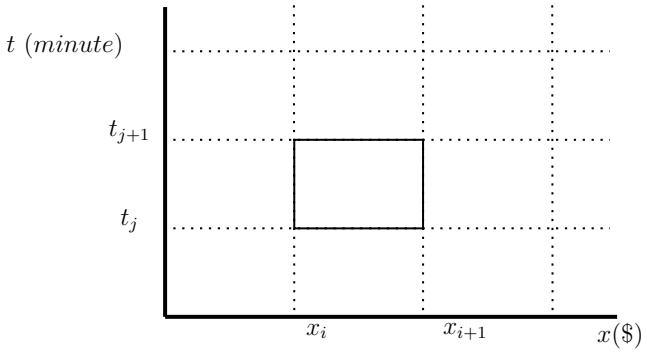


Figure 2.5: Space–Time Discretization

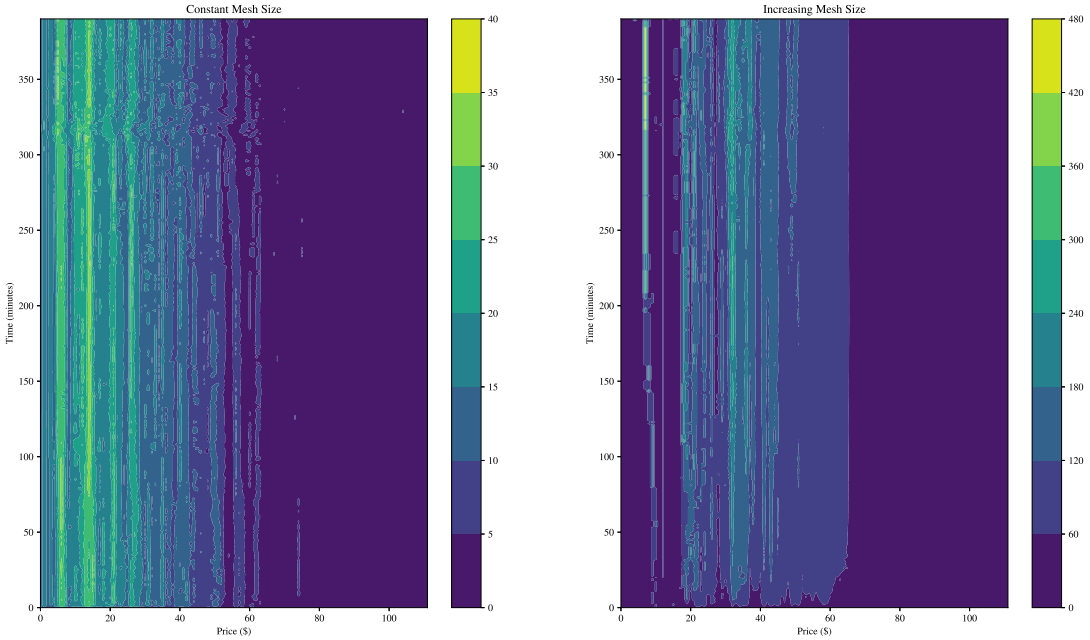


Figure 2.6: Density Heat Map

In Figure 2.6, we see heat maps for the two discretization methods; the horizontal axis shows the price x in dollars, and the vertical axis shows time t in minutes. When the mesh size is constant, the mass of stocks (or density) is concentrated in the first 40 price sections, while the density in the other mesh construction is greatest between \$20 and \$60. The heat map for the constant mesh size shows a sharp dip around the time of the flash crash, which isn't at all obvious in the case of the increasing mesh size. Thus, the constant mesh size was used.

2.3.3 Why Do We Need Macroscopic Variables?

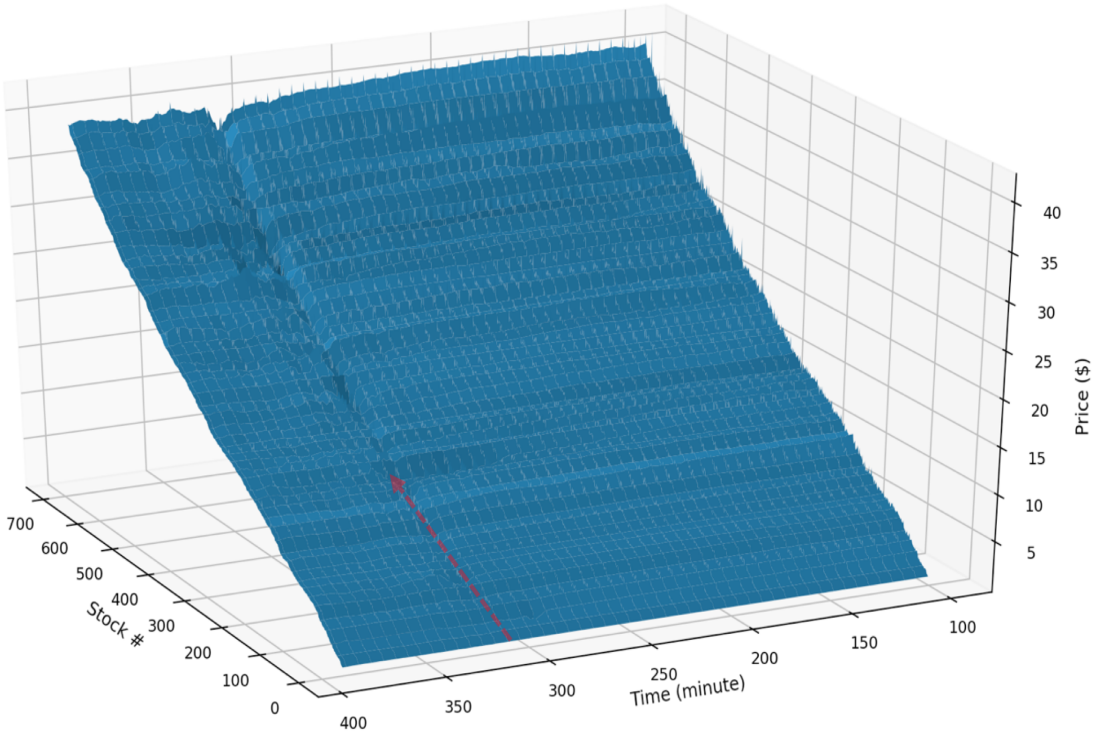


Figure 2.7: Movement of Stock Prices on May 6, 2010

Figure 2.7 shows a plot of the raw data, with a dip observed around the time of the flash crash. We observe that though the figure does capture some macroscopic behavior, such as the dip near the time of the crash, views such as the one in Figure 2.7 are helpful in illustrating the past behavior, but by their very nature they are not predictive of future such phenomena. Plots of the macroscopic variable velocity, for different price ranges are shown in Figure 2.8. This figure shows considerable variability and sensitivity around the time of the flash crash. The velocity plots suggest that equity markets were stable for a period of time, and then there was a shock that left the market unstable. This macroscopic variable is superior and promising, as it shows that the crash in fact started earlier.

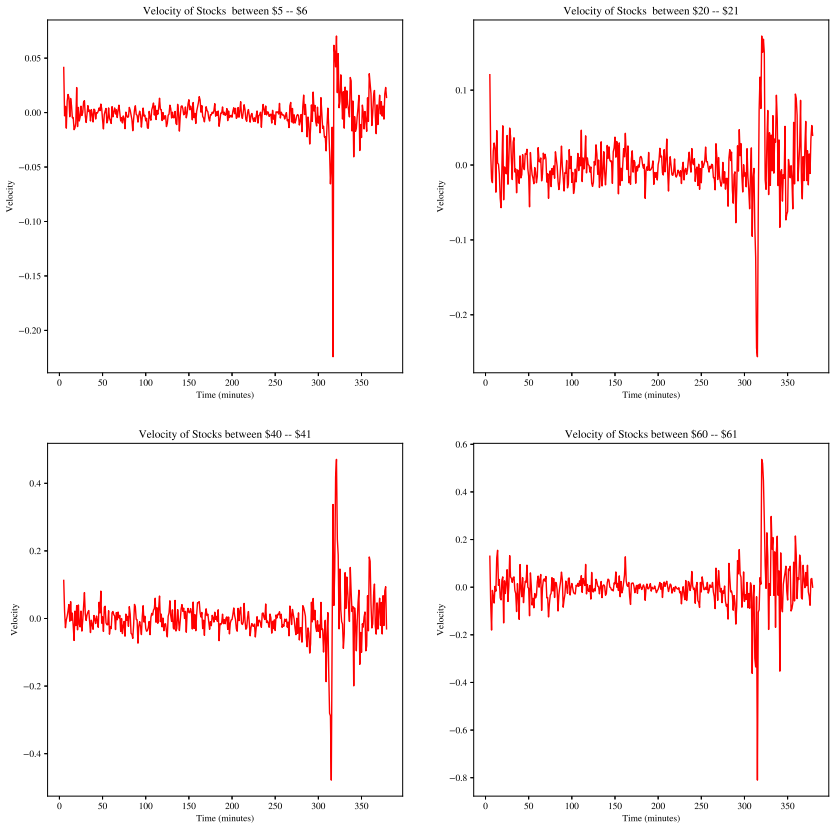


Figure 2.8: Velocity Plots

2.3.4 Fitting the Macroscopic Model

This section discusses in detail the process of fitting the right-hand sides of equations (2.25) and (2.26). From that process we attempted to understand whether the parameters exhibit unique behavior around the time of the flash crash, and analyzed them as sensors, in addition to the macroscopic measures.

The calibration of the parameters a, b, c, d, σ_1, and σ_2 at price x_i and time t_j was done by approximating the partial derivatives in equation (2.23) through (2.26) using a finite-

difference approximations to compute the values of the source terms $z_1(x_i, t_j)$ and $z_2(x_i, t_j)$, that is, the terms on the right-hand sides of equations (2.23) and (2.24), respectively (the expression on the RHS of a conservation law is normally called a source term).The parameter α in the pressure equation is set to 0.3 (which was found after some experimentation).

The time and space derivatives are approximated by forward and central difference approximations, respectively. The discretization intervals were chosen to be $\Delta t = 1$ and $\Delta x = 1$. Thus the system of stochastic partial differential equations was as follows:

$$\begin{aligned}\frac{\rho(x_i, t_{j+1}) - \rho(x_i, t_j)}{\Delta t} &+ \frac{\rho(x_{i+1}, t_j) - \rho(x_{i-1}, t_j)}{2\Delta x} v(x_i, t_j) \\ &+ \frac{v(x_{i+1}, t_j) - v(x_{i-1}, t_j)}{2\Delta x} \rho(x_i, t_j) = z_1(x_i, t_j)\end{aligned} \tag{2.28}$$

$$\begin{aligned}\frac{\rho(x_i, t_{j+1}) - \rho(x_i, t_j)}{\Delta t} v(x_i, t_j) &+ \frac{v(x_i, t_{j+1}) - v(x_i, t_j)}{\Delta t} \rho(x_i, t_j) \\ &+ (1 + 0.3) \frac{Q(x_{i+1}, t_j) - Q(x_{i-1}, t_j)}{2\Delta x} v(x_i, t_j) \\ &+ (1 + 0.3) \frac{v(x_{i+1}, t_j) - v(x_{i-1}, t_j)}{2\Delta x} Q(x_i, t_j) = z_2(x_i, t_j)\end{aligned} \tag{2.29}$$

$$z_1(x_i, t_j) = a(x_i, t_j) + b(x_i, t_j)\rho(x_i, t_j) + \sigma_1(x_i, t_j) \frac{\Delta W_1(x_{i+1} - x_{i-1}, t_{j+1} - t_j)}{2\Delta x \Delta t} \tag{2.30}$$

$$z_2(x_i, t_j) = c(x_i, t_j) + d(x_i, t_j)Q(x_i, t_j) + \sigma_2(x_i, t_j) \frac{\Delta W_2(x_{i+1} - x_{i-1}, t_{j+1} - t_j)}{2\Delta x \Delta t} \tag{2.31}$$

Equations (2.28) and (2.29) were fitted separately to obtain the values of the parameters on the RHS. Since the source terms $z_1(x_i, t_j)$ and $z_2(x_i, t_j)$ are linear functions of the density $\rho(x_i, t_j)$ and $Q(x_i, t_j)$, respectively, as shown in equations (2.30) and (2.31), we were able to apply a linear regression between $z_1(x_i, t_j)$ and $\rho(x_i, t_j)$ to estimate the parameters $a(x_i, t_j)$ and $b(x_i, t_j)$, and a linear regression between $z_2(x_i, t_j)$ and $Q(x_i, t_j)$ to estimate $c(x_i, t_j)$ and $d(x_i, t_j)$. The values of $\sigma_1(x_i, t_j)$ and $\sigma_2(x_i, t_j)$ were calculated from the equations

$$\sigma_1(x_i, t_j) = \sqrt{2\Delta x \Delta t \cdot E[\epsilon_1^2]} \tag{2.32}$$

$$\sigma_2(x_i, t_j) = \sqrt{2\Delta x \Delta t \cdot E[\epsilon_2^2]}, \tag{2.33}$$

where ϵ_1 and ϵ_2 are the residuals of the linear regressions in equations (2.30) and (2.31). The idea behind (2.30) and (2.31) is that the RHS of equation (2.31), for instance, can be rewritten as

$$z_2(x_i, t_j) = c(x_i, t_j) + d(x_i, t_j)Q(x_i, t_j) + \epsilon_2, \tag{2.34}$$

where ϵ_2 is an independent and identically normally distributed random variable with mean zero and variance $\frac{\sigma_2^2(x_i,t_j)}{2\Delta x \Delta t}$. Good model-fitting requires a sufficient number of samples, so we assumed that the parameters on the RHS of (2.31) are constant over a few cells. This is illustrated in Figure 2.9.

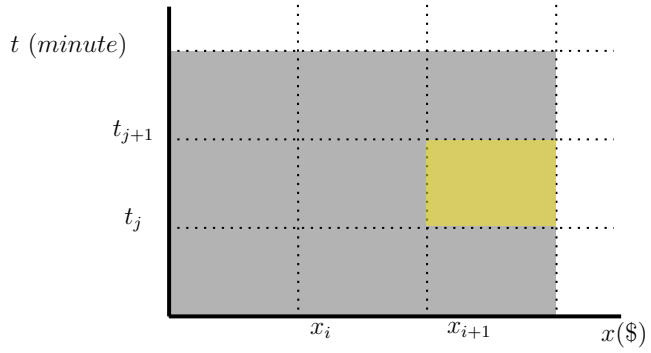


Figure 2.9: Space–Time Discretization

To calibrate the parameters for the yellow cell in Figure 2.9, we assumed that the parameters for all nine shaded cells in the figure are constant, and we used their cell information to fit the regression coefficients in (2.30) and (2.31). These values were assigned to the yellow cell. This procedure was used to estimate the parameters for each cell for the day of flash crash. We present and analyze only a selected number of representative samples from the

large number of results obtained.

Figures 2.10 through 2.15 show the fitted values of the parameters a, b, c, d, σ_1, and σ_2. The Jarque–Bera normality test was used to verify the normality assumption of the residuals [63]. We found that the residuals for more than 75% of the cells do not reject the normality hypothesis. The plots provide significant information about the day of the flash crash, which suggest that they might be useful as sensors for abnormal activity.

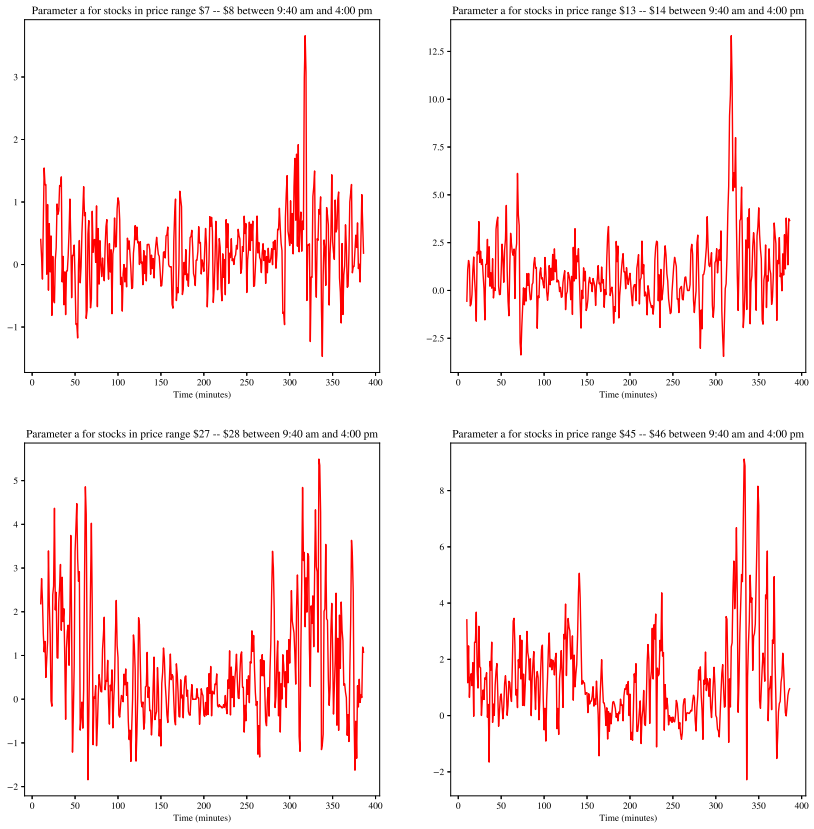


Figure 2.10: Four graphs of the fitted parameter a for May 6, 2010

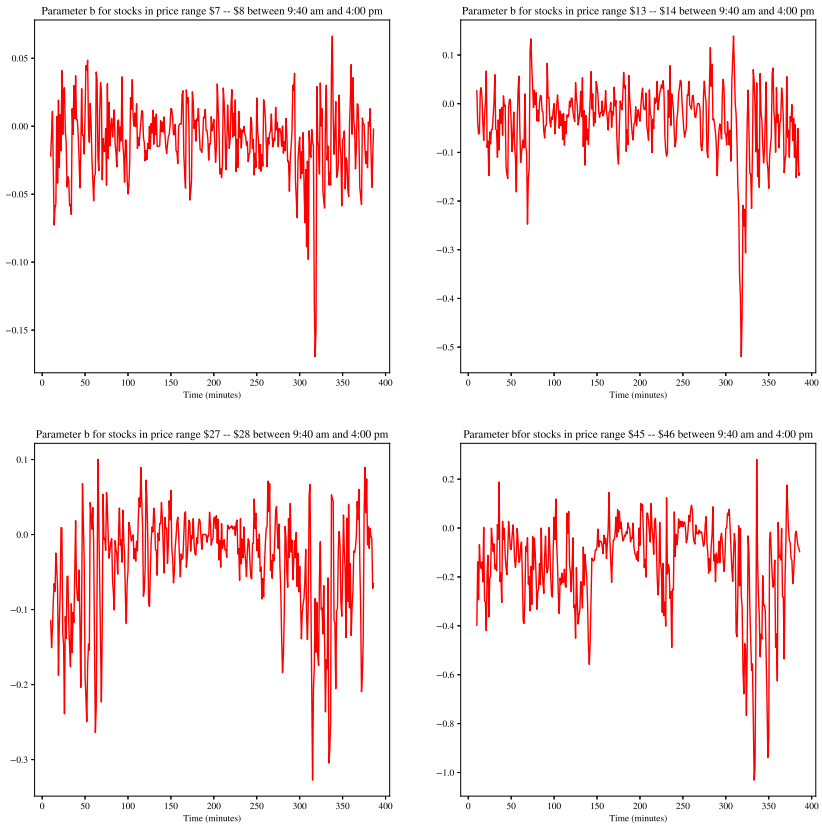


Figure 2.11: Four graphs of the fitted parameter b for May 6, 2010

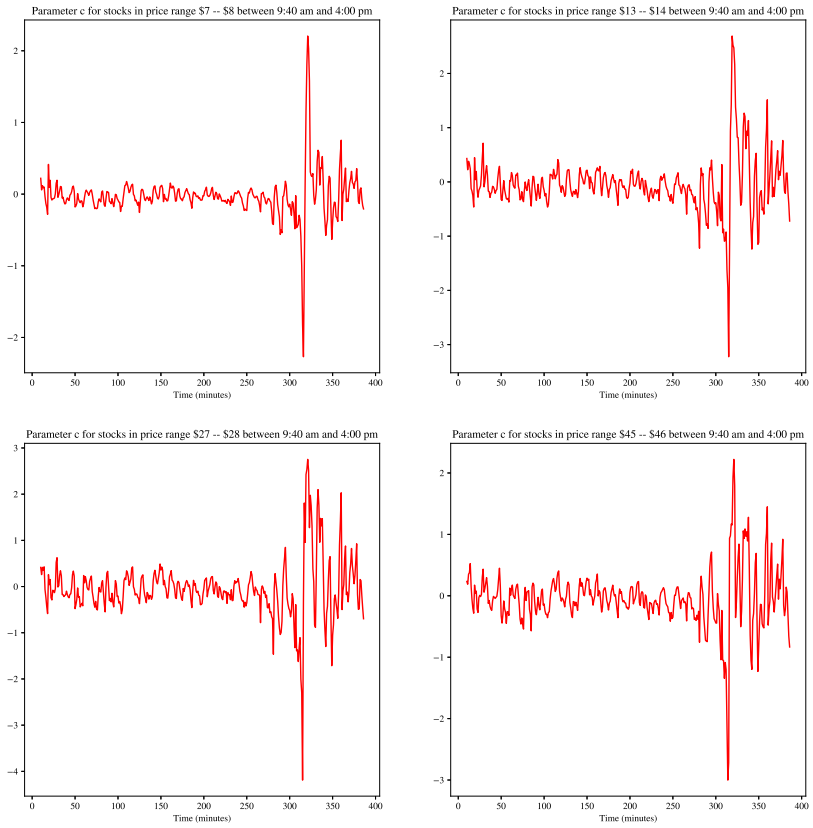


Figure 2.12: Four graphs of the fitted parameter c for May 6, 2010

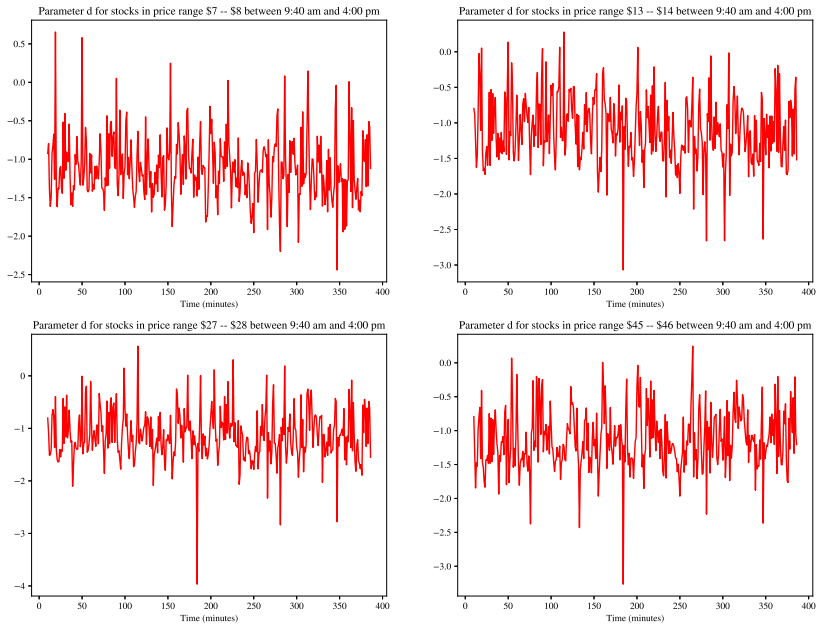


Figure 2.13: Four graphs of the fitted parameter d for May 6, 2010

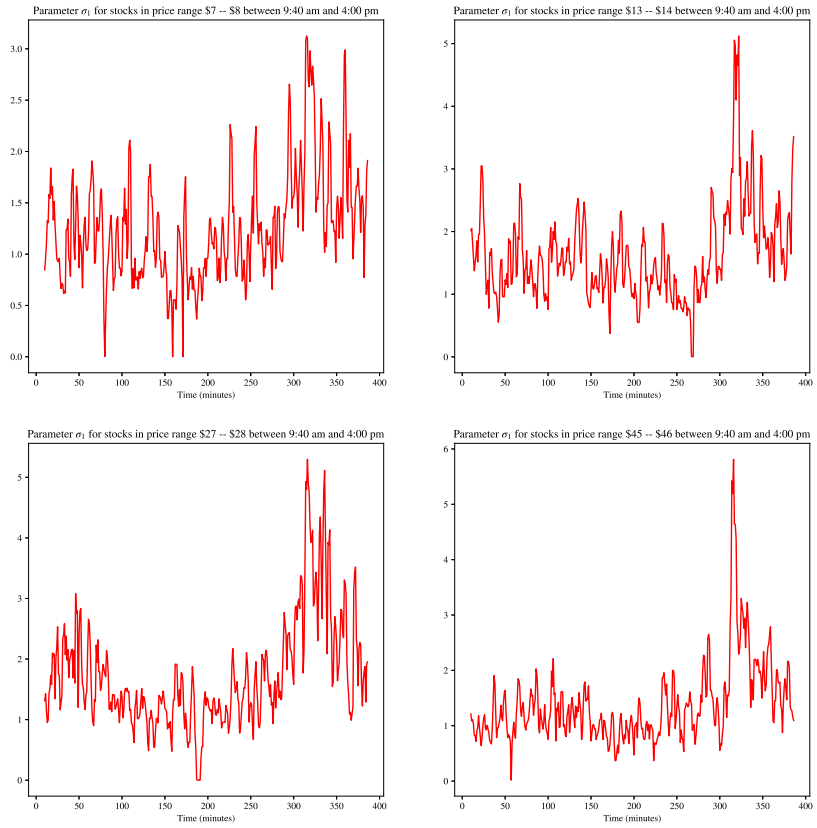


Figure 2.14: Four graphs of the fitted parameter σ_1 for May 6, 2010

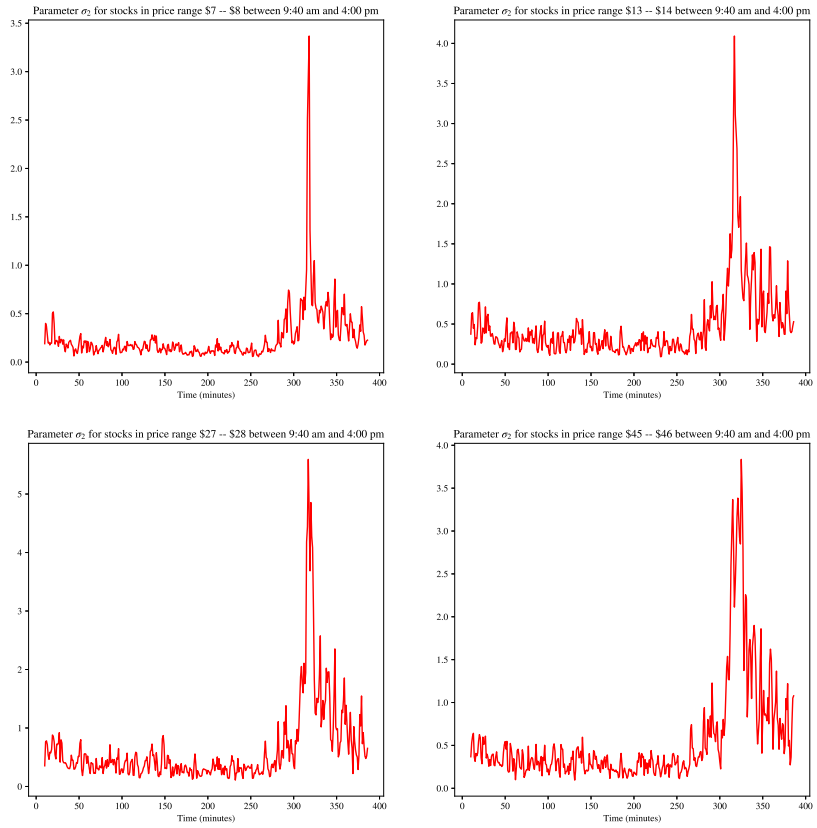


Figure 2.15: Four graphs of the fitted parameter σ_2 for May 6, 2010

2.3.5 Which Parameters to Monitor as Sensors?

Even though information is found in the parameters a, b and d, they are not consistent for all prices, whereas the parameters c, σ_1 and σ_2 showed consistency. A closer look at Figures 2.12, 2.14 and 2.15 demonstrates this. Figures 2.14 and 2.15 show that the volatility starts increasing prior to the announced flash crash time (302 minutes in the x-axis). We can also see that both σ_1 and σ_2 are higher after the flash crash than before. The parameter σ_2 appears to be a sensitive parameter, with a sharp increase during the flash crash. This initial study suggests that the parameters σ_1 and σ_2 as sensors may provide the required

on-line information for a device that can automatically inform the market of such activity. This finding is consistent with financial theory which states that volatility captures market conditions [105].

2.4 Conclusion

In this chapter, we introduced a new approach for the analysis of equity markets. By adapting the perspective of fluid dynamics, we introduced the concepts of density, velocity, flux, and pressure to capture the variability of the stock price in a market exchange. Borrowing the laws of conservation from physics, we developed the system of partial differential equations that describe the relationship among these variables. In validating the model, we added stochastic forcing terms to capture the noise generated by discretization of the model, data collection, and the approximation of the PDEs. To verify this model, we collected data from several markets and presented the results obtained from fitting and testing the model on a particular trading day. In the numerical results section, we analyzed the macroscopic measures with the aim of monitoring equity markets. We also showed how sensors can be generated by observing the behavior of the model parameters.

2.5 Technical Proofs

2.5.1 Proof of Theorem 2.2

The right eigenvectors corresponding to λ_1 and λ_2 are

$$r_1 = \begin{bmatrix} 1 \\ \\ \lambda_1 \end{bmatrix}, \; r_2 = \begin{bmatrix} 1 \\ \\ \lambda_2 \end{bmatrix}. \tag{2.35}$$

The eigenvalue λ_k is genuinely nonlinear if $\nabla\lambda_k(q) \cdot r_k(q) \neq 0$, which is the case here. The waves associated with this system of nonlinear partial differential equations admit only shocks

or rarefactions.

2.5.2 Proof of Theorem 2.3

Equation (2.19) can be rewritten as

$$\partial_t q + f(q)_x = 0, \tag{2.36}$$

where $f(q) := (Q, (1+\alpha)Qv)^T$. Introduce a test function $\phi \in C_0^1(\mathbb{R} \times \mathbb{R})$, where C_0^1 is the space of continuously differentiable functions with compact support. If we multiply equation (2.36) by $\phi(x,t)$ and then integrate over time and space, we get

$$\int_0^\infty \int_{-\infty}^\infty \Big[\phi q_t + \phi f(q)_x\Big] dx dt = 0 \tag{2.37}$$

Integrating (2.37) by parts yields

$$\int_0^\infty \int_{-\infty}^\infty \Big[\phi_t q + \phi_x f(q)\Big] dx dt = -\int_0^\infty \phi(x,0) q(x,0) dx \tag{2.38}$$

The boundary terms that arise from integrating by parts drop out since ϕ is assumed to have compact support. Therefore, q is a weak solution, since it satisfies (2.38) for all smooth functions ϕ. Note that q does not necessarily have to be differentiable to satisfy (2.38).

Chapter 3

Macroscopic Equity Markets Model: Towards Predicting Flash Crashes

3.1 Introduction

In the equity markets, the adaptation of new trading practices such as high frequency trading is increasing the number of market anomalies. New regimes are emerging, making human intervention to mitigate severe crises quite challenging [64]. One example of newly developed and unusual crashes (or extreme events) is the flash crash that occurred on May 6th, 2010 (hereafter known as, the flash crash day). On the flash crash day, the equity markets crashed suddenly, and the Dow Jones market exchange reported a drop of about 998 points. This drop in the index value is considered one of the most significant declines in Dow Jones history, which caused other financial indexes to also suffer [38].

Clearly, due to their complexity, understanding the dynamics of the financial markets is challenging. This complexity is attributable to the interactions between internal (or endogenous) and external (or exogenous) forces. While the internal forces are more significant to price changes than external ones [13] [66], the interactions between them are extremely important and can be measured by a quantity called reflexivity [45]. Reflexivty quantifies the ratio of stock price changes due to endogenous interactions to the overall interactions

(exogenous and endogenous). The author of [45] showed that the equity markets reflexivity increases during crashes which might provide an instrument to utilize in recognizing criticality (i.e., the tipping point of extreme events).

A primary concern in complex system studies is the predictability of extreme events. To be successful at this, new and innovative models are required. Attempts at modeling to predict extreme events in complex systems have lead to misleading results [110]. This deficiency motivated us to develop a non-conventional approach, inspired by nonlinear physics. Thus, in this research, we develop a model based on fluid flow which can predict disturbances and crashes. The use of physics in finance has a long history that goes back to the 19th century when Bachelier introduced the one-dimensional random walk (Brownian Motion) in finance [4]. After the discovery of Brownian Motion, the Geometric Brownian Motion (GBM) model became a central assumption to describe stock price movement [39]. GBM type models analyze stock prices using drift and volatility terms. Although this analysis is critical in microscopic studies, it ignores macroscopic information concerning stock interactions. Given the nonlinear and complex nature of stock interactions, major market events affect most market entities and are the product of the aggregate market disturbance and imbalance. If the goal is to monitor equity markets for early abnormality detection, it is more intuitive to do so macroscopically. The former approach, namely microscopic, is essential to study specific stocks. In contrast, the later or macroscopic approach gives more insight about the overall markets activity.

The vast majority of previous studies in physics and finance have investigated microscopic or mesoscopic (system interactions are described in detailed level between macroscopic and microscopic) models to detect crashes. In Chapter 2 [2], we proposed an analogy between fluid dynamics and stocks, while also deriving a Macroscopic Equity Markets (MEM) model. The MEM model assumes that stocks flow as fluid particles and describes macroscopic variables (i.e., density and flux) for its evaluation. In this model, stocks are aggregated and macroscopic variables such as density, and average velocity, are defined. The dynamics of

these macroscopic variables are described through a system of stochastic inhomogeneous partial differential equations which provide a realistic representation of the market dynamics. For more information about the model and related literature, we refer the reader to the paper [2].

In this chapter, we build on Chapter 2 by examining the properties of the MEM model. We show that during rapid movement of stock prices, the potential for shock wave formation exists. We propose real time integrative algorithms to solve the model dynamically. The proposed methods are adaptive and use numerical methods for solving a system of stochastic partial differential equations as well as stochastic filtering techniques. These algorithms are tested on datasets from normal and abnormal trading days, and the results suggest that unusual market conditions can be successfully identified by MEMs. After appropriate tuning, the proposed solution method is efficient and able to detect abnormalities.

The rest of the chapter is organized as follows. The properties of the MEM are investigated in section 2. Section 3 introduces the numerical method. Some stochastic filtering techniques are presented in section 4. The results are discussed in section 5. Section 6 concludes the chapter.

3.2 The Macroscopic Equity Markets Model

This section discusses the properties of the Macroscopic Equity Markets (MEM) model and reviews some background material that is necessary in our analysis. The MEM model was introduced in Chapter 2 [2] to connect the fluid-like properties of equity markets and external/unknown forces. The proposed model combines, in a natural way, external (exogenous) influences with internal (endogenous) dynamics. For simplicity, we use the following notations and definitions throughout this chapter. The space-time domain is discretized into small regions (called cells hereafter) of price (x) $\times$ time (t). The price domain (in \$) is denoted by x and the time domain (in minute) is denoted by t. In each cell, we adopt a fluid dynamics perspective to quantify stock price properties. Each stock is considered a single

mass unit in the price-time domain. To describe a stock's flow in the price-time domain, macroscopic variables velocity, density, flux and pressure are defined. The average velocity $v(x,t)$ is defined as the average change in stocks' prices at a given price x and time t. The density $\rho(x,t)$ is defined as the number of stocks per unit price x and unit time t. The flow of stocks is denoted by flux $Q(x,t) = \rho(x,t)v(x,t)$. The resulting MEM model is as follows:

$$\begin{aligned} \rho(x,t)_t + Q(x,t)_x &= a(x,t) + b(x,t)\rho(x,t) + \sigma_1(x,t)\frac{dW_1(x,t)}{dxdt}, \\ Q(x,t)_t + \Big((1+\alpha)Q(x,t)v(x,t)\Big)_x &= c(x,t) + d(x,t)Q(x,t) + \sigma_2(x,t)\frac{dW_2(x,t)}{dxdt}. \end{aligned} \tag{3.1}$$

The model above is a system of stochastic nonlinear Partial Differential Equations (PDEs). The right-hand side (called source or forcing term) of (3.1) represents a mean reversion process in which, after a disturbance, over time the density $\rho(x,t)$ and flux $Q(x,t)$ return to their historical average values. The Brownian sheet $W(x,t)$ is a Gaussian stochastic process that captures any disturbances not implicitly represented in the model. The source term is time-varying and is adapted to and captures the market conditions.

In fact, the MEM model has a plausible financial interpretation. The left hand side (LHS) of the equation (3.1) reflects the internal forces affecting a very small number of stocks. It captures the idiosyncratic or diversifiable risk. Internal forces and idiosyncratic risk are the drivers of stocks within the market, without the intervention of external forces. On the other hand, the right hand side (RHS) of (3.1) represents external forces and the systematic risk that is inherent in the whole market and can affect a large number of stocks. Noting that the MEM is a stochastic system of Partial Differential Equations (PDEs), the value of the parameter α of the LHS of equation (3.1), determines the type of PDEs. In Chapter 2, we demonstrated that there was a shock wave on the flash crash day indicating that the homogeneous MEM model should be a hyperbolic PDE (solutions of this type of PDEs are wave-like). Thus, the value of the parameter α should be chosen to preserve this property. In this chapter, we extend the previous work and show through extensive numerical results

that the the value of α is arbitrary and constant, but we pick α to be between zero and one for convenience. The sensitivity analysis on α is discussed in Section 3.5. Now we recall three theorems from [2]:

Theorem 3.1. *The system of partial differential equations in (3.1) is hyperbolic.*

Proof. Refer to [2] □

Theorem 3.2. *The system of partial differential equations in (3.1) admits only shocks or rarefaction waves.*

Proof. Refer to [2] □

Theorem 3.3. *The system of partial differential equations in (3.1) has a weak solution.*

Proof. Refer to [2] □

Remark 3.4. Equity markets exhibit wave motion phenomenon in which information about stocks (particles), flow at a speed determined by the market conditions. A forward flowing shock carries information from lower priced stocks to higher priced ones, and a backward shock carries information from the higher priced to lower priced stocks. The phenomenon of waves is observed in many applications, including traveling sound waves in physics and traveling velocity waves in traffic flows [127].

Though the nonlinearity in (3.1) poses computational challenges [40], it is useful for modeling abnormal behaviors of equity markets, especially when abrupt changes in stock prices occur. With an appropriate choice of the parameter α, the model admits shock or rarefaction waves (Theorem 3.2). A shock is defined as a discontinuity and/or jump in density or velocity. In Chapter 2, we showed that shock waves are observed in equity markets, but have relatively low speeds, and thus can be missed. On the other hand, rarefaction waves are more apparent in physical systems. In this research, we describe rarefaction waves as gradual increase or decrease in the macroscopic variables. Indeed, such behavior might be

attributed to the recovery process of markets. Explicit conditions for forming shocks and rarefaction waves are discussed in Section 3.2.2.

Definition 3.5. A financial shock is defined as a discontinuity or jump in density or velocity of stocks at price x and time t.

Definition 3.6. A financial rarefaction wave is a gradual increase or decrease in the macroscopic variables (i.e., density or velocity).

In the next section, we establish some results about systems of generic nonlinear hyperbolic PDEs. These results are needed in the subsequent sections where we derive the general solutions of (3.1). In Section 3.2.2, the structure of the model is analyzed with the RHS zero. The resulting system is referred to as homogeneous. The analysis of the homogeneous system gives insights into the behavior of the macroscopic variables during periods of abrupt changes.

3.2.1 Preliminaries

This section reviews results about hyperbolic conservation laws (for a comprehensive overview about PDEs and numerical methods, see [40] and [82]). These results are needed to establish the theoretical analysis of shock and rarefaction waves in equity markets. We view the flow of stocks as fluid, and the Eulerian framework, in which fluid properties are represented as a function of position x and time t. A system of conservation laws is a classical example of hyperbolic PDEs and takes the integral form for all (x, t)

$$\frac{d}{dt}\int_{x_1}^{x_2} q(x,t)dx + \int_{x_1}^{x_2} \frac{\partial}{\partial x} F(q(x,t))dx = \int_{x_1}^{x_2} S(x,t)dx, \tag{3.2}$$

where $q \in \mathbb{R}^m$ is the vector of the conserved quantities, $F \in \mathbb{R}^m$ is the vector of the fluxes of the conserved quantities, $S \in \mathbb{R}^m$ is the vector of source terms. The equation (3.2) states that the rate of change of a conserved quantity q over a fixed section (x_1, x_2) is due to the effects of the net flux F through the endpoints of that section and the flow of external sources

S. When q and F are smooth functions, the system (3.2) can be written equivalently in the differential form as for all (x, t)

$$q_t(x,t) + F(q(x,t))_x = S(x,t) \tag{3.3}$$

or

$$q_t(x,t) + F'(q(x,t))q_x(x,t) = S(x,t) \tag{3.4}$$

The MEM model (3.1) is in the differential form of (3.2) when $m = 2$ and $x \in \mathbb{R}^+$. This form is not valid when q is not smooth (i.e., discontinuous). In particular, these equations become invalid when a discontinuity propagates due to the nonlinearity or jumps in the initial conditions (i.e., Riemann Problems). In that case, the integral form should be used instead to determine the proprieties of the discontinuity, while the differential form is confined to smooth regions.

Definition 3.7. Many problems arise from one-dimensional Riemann Problems. When $q_L \neq q_R$, these problems take the form

$$q_t(x,t) + F(q(x,t))_x = 0 \text{ for } t > 0 \text{ and } x \in \mathbb{R} \tag{3.5}$$

$$q(x,0) = \begin{cases} q_L, & x < 0 \\ q_R, & x > 0 \end{cases} \tag{3.6}$$

Lemma 3.8. *[82] The shock speed propagates with speed s and can be determined in terms of the left q_L and right q_R states by satisfying the Rankine-Hugoniot jump condition which is given by*

$$(F_R - F_L) = (q_R - q_L)s, \tag{3.7}$$

where s is the speed of the discontinuity.

Proof. See [82]. □

If a shock propagates with speed s between q_L and q_R on each side of the shock, the Rankine-Hugoniot jump condition must hold. This condition is, in fact, derived from the integral form of the conservation laws. This reinforces the importance of relaying on the integral form for determining discontinuous solutions. In order to obtain a unique solution for q associated with propagating discontinuities, it is important to introduce the **Lax Admissibility Condition**. This condition is applied to discontinuous solutions to determine which jumps are allowed.

Lemma 3.9. *[78] Suppose that the characteristic speeds are real and either linear or nonlinear degenerate. Assume that those characteristic speeds are distinct as well. If the characteristics speeds λ_i are ordered from least to greatest, then a discontinuity that propagates with speed s is a shock, if and only, if there exists an index i such that the following inequalities hold*

$$(\lambda_i)_R < s < (\lambda_i)_L \tag{3.8}$$

Proof. See [78]. □

Shocks are discontinuous weak solutions that satisfy the entropy condition above. The Riemann Problem in the system of nonlinear hyperbolic conservation laws admit rarefaction waves (which are smooth and correct solutions) when $(\lambda_i)_L < (\lambda_i)_R$ for any i.

3.2.2 Theoretical Analysis: Shocks and Rarefaction Waves

The purpose of this section is to introduce the conditions of which a pair of states (either density or flux of stocks) can be connected by a shock (by determining the Hugoniot Locus) or rarefaction wave (by determining the Integral Curves). The shock speed and direction are determined as well. The structural properties of the MEM model are established on the homogeneous system only since the forcing term does not affect the eigen-structure of the system of equations.

Consider the homogeneous MEM model

$$\rho_t + Q_x = 0. \tag{3.9}$$

$$Q_t + \Big((1+\alpha)Qv\Big)_x = 0. \tag{3.10}$$

Equations (3.9) and (3.10) can be rewritten in the compact form

$$q_t + F(q)_x = 0, \tag{3.11}$$

or

$$q_t + F'(q)q_x = 0, \tag{3.12}$$

where $F(q) = (Q, (1+\alpha)Qv)^\top$, $q = (\rho, Q)^\top$ is a vector [1] with elements that correspond to the state variables, density and flux, and

$$F'(q) = \begin{bmatrix} \frac{\partial F_1}{\partial q_1} & \frac{\partial F_1}{\partial q_2} \\ \frac{\partial F_2}{\partial q_1} & \frac{\partial F_2}{\partial q_2} \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ -(1+\alpha)\frac{q_2^2}{q_1^2} & 2(1+\alpha)\frac{q_2}{q_1} \end{bmatrix}. \tag{3.13}$$

The eigenvalues of $F'(q)$ (called information speed) are

$$\lambda_{1,2} = (1+\alpha \pm \sqrt{\alpha^2+\alpha})\frac{q_2}{q_1}. \tag{3.14}$$

The eigenvectors associated with λ_1 and λ_2 are

$$r_1 = \begin{bmatrix} 1 \\ \lambda_1 \end{bmatrix},\ r_2 = \begin{bmatrix} 1 \\ \lambda_2 \end{bmatrix}. \tag{3.15}$$

The system (3.11) is strictly hyperbolic (Theorem 3.1) when the eigenvalues in (3.14)

[1] For simplicity, we denote $q_1 = \rho$ and $q_2 = Q$.

are distinct and real; that is, when $\alpha > 0$ or $\alpha < -1$, and $\frac{q_2}{q_1} \neq 0$. Hereafter, the value of α is assumed to be **constant** and **between zero and one**. In Section 3.5, we provide more analysis on the choice of α.

As previously noted, the nonlinear system admits rarefaction and/or shock waves. A valid shock is a solution that satisfies the Rankin-Hugoniot condition of Lemma 3.8 and the Entropy condition of Lemma 3.9. Assume that a shock or discontinuity propagates with a speed s and has values q_m and q_* on each side of the discontinuity. The following condition must hold

$$F(q_m) - F(q_*) = (q_m - q_*)s, \tag{3.16}$$

To determine the set of states q_m that can be connected by a discontinuity through q_*, we use the Rankin-Hugoniot jump condition with an initial state q_* that can be q_L or q_R

$$Q_m - Q_* = s(\rho_m - \rho_*) \tag{3.17}$$

$$(1+\alpha)\Big(\frac{Q_m^2}{\rho_m}\Big) - (1+\alpha)\Big(\frac{Q_*^2}{\rho_*}\Big) = s(Q_m - Q_*) \tag{3.18}$$

The system in (3.17) and (3.18) leads to two different families of solutions. The system can be solved for Q_m and s in terms of ρ_m. From (3.17), we obtain

$$s = \frac{(Q_m - Q_*)}{(\rho_m - \rho_*)}. \tag{3.19}$$

Replacing s in (3.18) by (3.17), and after a few steps (the full derivation is given in the Appendix) we obtain

$$Q_m = Q_* \frac{1 \pm (\rho_m - \rho_*)\sqrt{\frac{\alpha^2+\alpha}{\rho_m \rho_*}}}{1 - \frac{(\rho_m - \rho_*)}{\rho_m}(1+\alpha)}. \tag{3.20}$$

There are two families of curves that connect a left state Q_* and a right state Q_m. The

sign in (3.20) is determined when we discuss the conditions of the physical (valid) jumps. In order to determine the density ρ_R, we may simply equate the flux equation as in (3.20) across the shock from left flux (Q_*) to right flux (Q_m) and right flux (Q_m) to left flux (Q_*). That must be true since the middle state should be the same whether we jump from the right flux or left flux.

The Riemann problem for (3.11) around $x = 0$ has the following jump conditions

$$q(x, t = 0) = \begin{cases} q_L = (\rho_l, Q_l), \text{ if } x < 0 \\ q_R = (\rho_r, Q_r), \text{ if } x > 0 \end{cases} \tag{3.21}$$

The previous problem states that the initial condition for the system in (3.11) has a jump around a price $x = 0$ with a left state q_l and right state q_r. The value $x = 0$ is an arbitrary price location and it could be any price x with different left and right states. The problem in (3.21) is solved by shock and/or rarefaction waves, depending on the values of left and right states. To begin, we analyze the shock waves by pointing out that there are two distinct types of shock waves for (3.11); we call them 1-shock and 2-shock. The Lax entropy condition (3.9) for 1-shock corresponding to the first eigenvalue λ_1 states that

$$\lambda_1(q_R) < s < \lambda_1(q_L), \tag{3.22}$$

where s is the shock speed and is given by the Rankin-Hugoniot condition (3.16). This yields

$$(1 + \alpha - \sqrt{\alpha^2 + \alpha}) \frac{Q_R}{\rho_R} < s < (1 + \alpha - \sqrt{\alpha^2 + \alpha}) \frac{Q_L}{\rho_L}, \tag{3.23}$$

where $s = \frac{Q_R - Q_L}{\rho_R - \rho_L}$.

The entropy condition is a mathematical inequality that ensures that only relevant solutions are considered as shocks. This relevancy in the equity markets is outlined as follows. The financial shock occurs when stock velocity (average price change) jumps to a lower value. This can be interrupted as shocks originating between fast and slow stocks, moving faster

than low velocity stocks but slower than fast velocity stocks. The velocity here coincides with the information speed as represented by $\lambda_{1,2}$. In the next few theorems, we highlight insights about market shocks.

Theorem 3.10. *The Lax Entropy Condition (3.23) is satisfied if $Q_R > Q_L$, and $\rho_R > \rho_L$ (when $Q_L > 0$) or $\rho_R < \rho_L$ (when $Q_L < 0$). The curve S1 is expressed as*

$$S1 : Q_R = Q_L \frac{1 - (\rho_R - \rho_L)\sqrt{\frac{\alpha^2+\alpha}{\rho_R \rho_L}}}{1 - \frac{(\rho_R - \rho_L)}{\rho_R}(1+\alpha)}, \tag{3.24}$$

the corresponding shock speed

$$s_1 = Q_L \frac{\frac{1+\alpha}{\rho_R} - \sqrt{\frac{\alpha^2+\alpha}{\rho_R \rho_L}}}{1 - \frac{(\rho_R - \rho_L)}{\rho_R}(1+\alpha)}, \tag{3.25}$$

Similarly, the Lax entropy condition for 2-shock is

$$(1 + \alpha + \sqrt{\alpha^2+\alpha})\,\frac{Q_R}{\rho_R} < s < (1 + \alpha + \sqrt{\alpha^2+\alpha})\,\frac{Q_L}{\rho_L}, \tag{3.26}$$

Theorem 3.11. *The Lax Entropy Condition (3.26) is satisfied if $Q_R < Q_L$, and $\rho_R < \rho_L$ (when $Q_L > 0$) or $\rho_R > \rho_L$ (when $Q_L < 0$). The curve S2 is expressed as*

$$S2 : Q_R = Q_L \frac{1 + (\rho_R - \rho_L)\sqrt{\frac{\alpha^2+\alpha}{\rho_R \rho_L}}}{1 - \frac{(\rho_R - \rho_L)}{\rho_R}(1+\alpha)}, \tag{3.27}$$

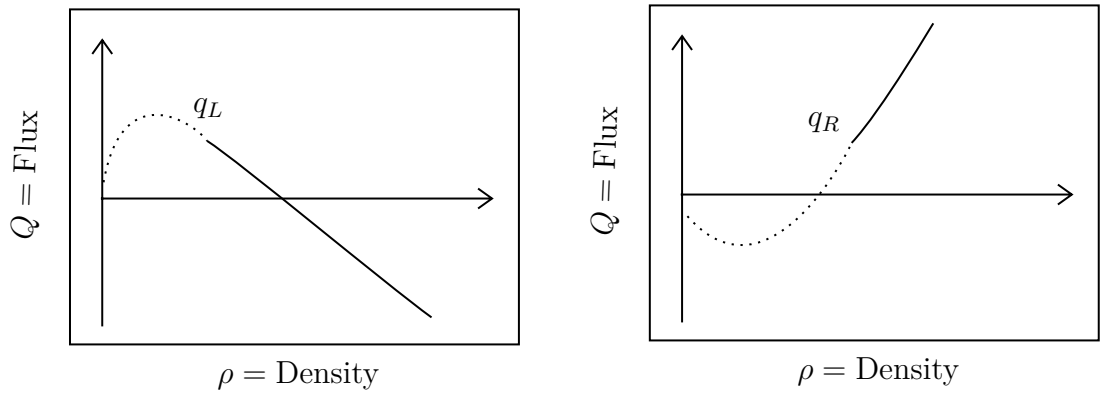
the corresponding shock speed

$$s_2 = Q_L \frac{\frac{1+\alpha}{\rho_R} + \sqrt{\frac{\alpha^2+\alpha}{\rho_R \rho_L}}}{1 - \frac{(\rho_R - \rho_L)}{\rho_R}(1+\alpha)}, \tag{3.28}$$

It is interesting to see the jump between q_L and q_R in the phase plane for the system of two equations. This means the plane of $q_1 - q_2$ (i.e., $\rho - Q$ plane). Each point in the

plane represents a vector $q(x,t) = (\rho(x,t), Q(x,t))^\top$. To be more specific, the left and right states, q_L and q_R, are points in this plane. From the Rankin-Hugoniot condition, we see that a jump or discontinuity between q_L and q_R propagates as a single jump only if $q_R - q_L$ is an eigenvector of $F(q)$. This implies that a line connecting a jump from q_L to q_R must be parallel to either r_1 or r_2. These lines are called Hugoniot Locus. Figure 3.1 shows the Hugoniot Locus of the MEM model in the $\rho - Q$ plane. This figure represents a case when $Q_L > 0$. The other case is omitted here because it can be drawn similarly.

A right state q_R is connected to a left state q_L by 1-shock if it lies on the S1 curve, passing through q_L. Similarly, the right state is connected to the left state by 2-shock if it lies on the S2 curve passing through q_L. If the right state is not connected to a left state by a shock curve, a middle state q_m is determined to connect q_L to q_R by two types of curves.



(a) $S1$: Shock-1 (b) $S2$: Shock-2

Figure 3.1: Hugoniot Locus: Solid lines indicate the Entropy-Satisfying Jumps

Proposition 3.12. *The conditions that determine the direction of the first and second shocks are as follows:*

- *If* $\frac{\rho_L}{\rho_R} < \frac{\alpha}{1+\alpha}$ *and* $Q_L > 0$*, then we obtain that* $s_1 > 0 > s_2$*.*
- *If*$\frac{\rho_L}{\rho_R} < \frac{\alpha}{1+\alpha}$ *and* $Q_L < 0$*, then we obtain that* $s_2 > 0 > s_1$*.*
- *If* $\frac{\rho_L}{\rho_R} > \frac{\alpha}{1+\alpha}$ *and* $Q_L > 0$*, then we obtain that* $s_2 > s_1 > 0$*.*

- *If* $\frac{\rho_L}{\rho_R} > \frac{\alpha}{1+\alpha}$ *and* $Q_L < 0$*, then we obtain that* $s_2 < s_1 < 0$.

When a shock is formed, its direction depends on the initial right and left states (i.e., density and flux). In the markets, a positive shock indicates that when a shock starts it moves from low to high price stocks, while the negative shock moves in the other direction. This insight was validated in Chapter 2, in which a shock was observed to be positive on the flash crash day.

The Riemann problem admits rarefaction waves. The rarefaction occurs, theoretically speaking, when prices gradually recover. That is, when the left information speed $(\lambda_i)_L$ is slower than the right one $(\lambda_i)_R$ for $i = 1, 2$, rarefaction waves, called 1-Rarefaction and 2-Rarefaction, exist. To visualize this solution, consider a smooth recovery of stocks after a shock. Such smooth transition connects two states, q_L and q_R. If we consider the MEM model in (3.12). The solution of the system (3.12) is scaling-invariant [81]. That is to say, $q(x,t)$ depends on (x,t) in the form $\zeta = x/t$. Therefore, a rarefaction wave solution of the Riemann problem (3.21) takes the form

$$q(x,t) = \begin{cases} q_L = (\rho_l, Q_l), \text{ if } x \leq \zeta_1 t \\ \tilde{q}(x/t), \text{ if } \zeta_1 t < x < \zeta_2 t \\ q_R = (\rho_r, Q_r), \text{ if } x \geq \zeta_2 t. \end{cases} \tag{3.29}$$

where $q_L = \tilde{q}(\zeta_1), q_R = \tilde{q}(\zeta_2)$, and $\tilde{q}(x/t)$ is a smooth function. There are two curves composed of points q_R that can be connected to q_L. The left state q_L is connected to the right state q_R by a rarefaction wave. The two curves are a subset of the integral curves of the vector field $r_i(q)$ [80]. More specifically, rarefaction waves lie along integral curves. To see this, we differentiate $q(x,t) = \tilde{q}(x/t)$ with respect to time and price x to obtain

$$q_t(x,t) = -\frac{x}{t^2}\tilde{q}'(x/t) \tag{3.30}$$

$$q_x(x,t) = \frac{1}{t}\tilde{q}'(x/t) \tag{3.31}$$

substituting the derivatives in (3.12), we obtain

$$-\frac{x}{t^2}\tilde{q}'(x/t) + \frac{1}{t}F'(x/t)\tilde{q}'(x/t) = 0 \tag{3.32}$$

yielding the following relationship:

$$F'(\tilde{q}(\zeta))\tilde{q}'(\zeta) = \zeta\tilde{q}'(\zeta), \tag{3.33}$$

where $\zeta = x/t$. Equation (3.33) says that $\tilde{q}'(\zeta)$ is proportional to an eigenvector $r_i(\tilde{q}(\zeta))$ of $F'(\tilde{q}(\zeta))$. That is to say,

$$\tilde{q}'(\zeta) = \mu(\zeta)r_i(\tilde{q}), \tag{3.34}$$

for some $i = 1, 2$. Hence, the values of $\tilde{q}(\zeta)$ lie along an integral curve of $r_i(\tilde{q})$. The left and right states, $\tilde{q}(\zeta_1)$ and $\tilde{q}(\zeta_2)$, must lie on the same integral curve. This condition is necessary for a rarefaction wave to exist between q_R and q_L. Another condition for the existence of a rarefaction wave connecting q_L and q_R is that our parameterization $\zeta = x/t$ should be monotonically increasing as we move from the left state q_L to the right state q_R. From (3.33), we see that ζ is the eigenvalue of $F'(\tilde{q}(\zeta))$. In particular,

$$\zeta = \lambda_i(\tilde{q}(\zeta)). \tag{3.35}$$

The phase curves and conditions in which rarefaction waves formed are outlined below in two theorems.

Theorem 3.13. *The 1-rarefaction wave curve is described as follows:*

$$R1 : Q_R = Q_L\Big(\frac{\rho_R}{\rho_L}\Big)^{1+\alpha-\sqrt{\alpha^2+\alpha}}, \tag{3.36}$$

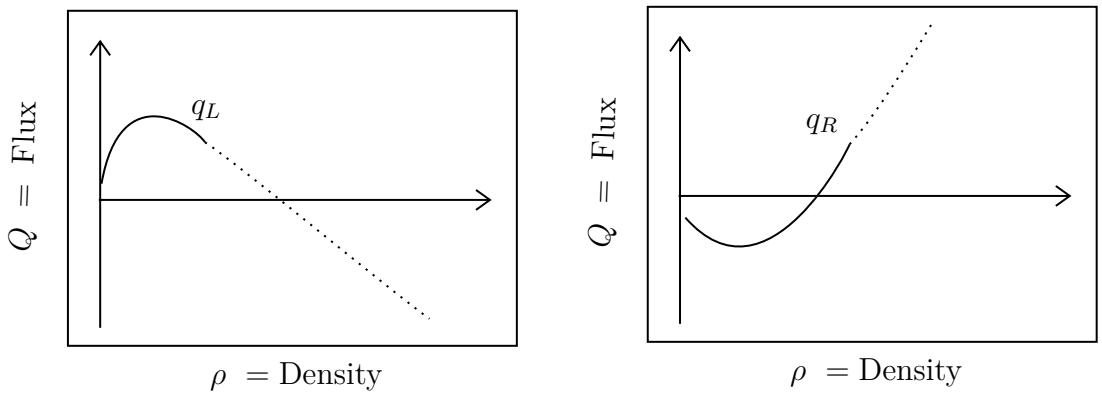
also $\rho_R < \rho_L$ *when* $Q_L > 0$ *and* $\rho_R > \rho_L$ *when* $Q_L < 0$.

Theorem 3.14. *The 2-rarefaction wave curve is described as follows:*

$$R2 : Q_R = Q_L\Big(\frac{\rho_R}{\rho_L}\Big)^{1+\alpha+\sqrt{\alpha^2+\alpha}}, \tag{3.37}$$

also $\rho_R > \rho_L$ *when* $Q_L > 0$ *and* $\rho_R < \rho_L$ *when* $Q_L < 0$.

The Integral Curves for the MEM model (for the case when $Q_L > 0$) is shown in Figure 3.2. The other case is omitted here because it can be drawn similarly.



$(a)\, R1$: Rarefaction-1 $(b)\, R2$: Rarefaction-2

Figure 3.2: Integral Curves: Solid lines indicate the correct rarefaction waves

A right state q_R is connected to a left state q_L by a 1-Rarefaction wave if it lies on the R1 curve passing through q_L. Similarly, the right state is connected to the left state by a 2-Rarefaction wave if it lies on the R2 curve passing through q_L. If the right state is not connected to a left state by a rarefaction curve, a middle state q_m is determined to connect q_L to q_R by two types of curves.

In this section, we examined the properties of the model and showed that shock and rarefaction waves can be formed, but the system is yet to be solved and show how the density and velocity of stocks evolve over time. Indeed, that is essential for predicting market conditions. A linear version of the system of PDEs in (3.1) is usually solved by the method of characteristics, in which characteristic lines are established and the solutions are constant along them. With the formation of discontinuities, computational difficulties arise. In this case, approximation methods are preferred in many practical applications to approximate discontinuous solutions. Difference methods, which approximate derivatives by finite differences, are expected to break down around such solutions. To overcome this issue, finite volume methods are employed in this chapter. This method works with the integral form of the system of equations. The finite volume methods do not approximate derivatives point-wise, but the domain is initially discretized into grid cells and the total integral (average changes) of the variables are approximated over each cell. As we will see in Section 3.3, the integrals are only modified by the flux and/or forcing term through the edges of cells. The finite volume numerical methods obtain a conservative form. That is to say, it provides the same property of the exact solution. The importance of the conservative form is that the correct shock speeds are captured, while it might not be the case when a non-conservative form is used.

Given the aforementioned challenges and nature of the problem (being nonlinear and stochastic), we propose numerical approaches based on finite volume methods to solve the system in real-time. To do so, we decompose the system into three parts, depending on the type of forcing term, namely: homogeneous, deterministic inhomogeneous, and stochastic inhomogeneous. Numerical methods are proposed for homogeneous and deterministic inhomogeneous systems. The stochastic inhomogeneous system is tackled through stochastic filtering techniques from estimation theory.

3.3 Methodology

This section proposes various numerical techniques to solve the MEM model. The solution methodology consists of two main components: numerical methods and stochastic filtering techniques. We purpose an integrative numerical method for the homogeneous model and the deterministic inhomogeneous one. The stochastic system is discussed in Section 4.

3.3.1 Homogeneous MEM Model

We adapt the finite volume method to solve and evaluate the model. The finite volume method discretizes the spatial domain (i.e., the price domain) being divided into grid cells, and approximates the integral of the vector of macroscopic variables, say q, over each of the resulting cells. The information in each cell is updated at each time step using an approximation to the flux. We define a grid cell C_i by an interval $(x_{i+\frac{1}{2}}, x_{i-\frac{1}{2}})$ as shown in Figure 3.3.

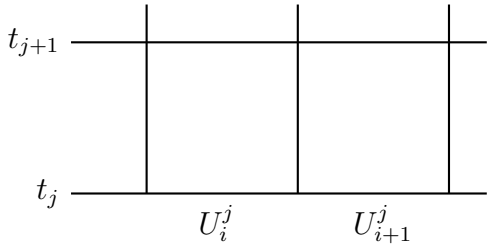


Figure 3.3: Grid Cells C_i and C_{i+1}

where U_i^j is approximately defined as the average of the integral of q over the grid cell i at time t_j

$$U_i^j \approx \frac{1}{(x_{i+\frac{1}{2}} - x_{i-\frac{1}{2}})} \int_{x_{i-\frac{1}{2}}}^{x_{i+\frac{1}{2}}} q(x, t_j)dx = \frac{1}{\Delta x} \int_{x_{i-\frac{1}{2}}}^{x_{i+\frac{1}{2}}} q(x, t_j)dx. \tag{3.38}$$

The finite volume method works with cell averages to ensure that the numerical method is conservative in order to accurately capture discontinuous solutions [82]. The numerical

form of the MEM model in (3.11) is given by

$$U_i^{j+1} = U_i^j - \frac{\Delta t}{\Delta x}(F^j_{i+\frac{1}{2}} - F^j_{i-\frac{1}{2}}), \tag{3.39}$$

where equation (3.39) determines the time evolution of the cell average U_i. The value $F^n_{i+\frac{1}{2}}$ approximates the average flux through the interface $x_{i+\frac{1}{2}}$, i.e.,

$$F^j_{i+\frac{1}{2}} \approx \frac{1}{\Delta t}\int_{t_j}^{t_{j+1}} F(q(x_{i+\frac{1}{2}}, t))dt. \tag{3.40}$$

Denote the numerical flux by

$$F^j_{i+\frac{1}{2}} = \mathbb{F}(U_i^j, U_{i+1}^j). \tag{3.41}$$

Equation (3.41) says that the flux value can be obtained based on the cell averages on either side of the interface. Given (3.41), the numerical form of (3.39) becomes

$$U_i^{j+1} = U_i^j - \frac{\Delta t}{\Delta x}(\mathbb{F}(U_i^j, U_{i+1}^j) - \mathbb{F}(U_{i-1}^j, U_i^j)). \tag{3.42}$$

The finite volume method presented in (3.42) is stable and will converge to the solution of the differential equations if the grid is refined to satisfy the CFL condition [30]. This condition ensures that the numerical domain contains the mathematical or physical domain. The CFL condition for hyperbolic PDEs is given by

$$\frac{\Delta x}{\Delta t}\max\{|\ \lambda_1\ |, |\ \lambda_2\ |\} \leq 1. \tag{3.43}$$

There are various methods to approximate $\mathbb{F}(\cdot)$, but one popular approach is Godunov's method [51]. This method attempts to solve the Riemann problem at each interface. By employing Godunov's method, the numerical flux at the interface $x_{i+\frac{1}{2}}$ can be expressed as

$$F^j_{i+\frac{1}{2}} = F(q(U^j_i, U^j_{i+1})), \tag{3.44}$$

where $q(U^j_i, U^j_{i+1})$ is the solution of the Riemann problem between U^j_{i+1} and U^j_i. We adapt the Godunov's method to approximate the flux of stocks through the interfaces. In that context, the Riemann problem in (3.44) can be potentially solved by using the complete description of the solution as introduced in Section 3.2. This process is quite expensive and requires the entire structure of the Riemann problem. However, in order to compute $q(U^j_i, U^j_{i+1})$, for instance, its value can be roughly assumed to be one of the states connecting U^j_i to U^j_{i+1} through a number of jumps (shock or rarefaction waves). The intermediate state can be viewed graphically as one of the intersections of the Hugoniot Locus and/or Integral Curves [80]. To reduce the computational effort, we use approximate Riemann solvers to replace the full Riemann structure . Approximate Riemann solvers have, not only shown to be as good as the exact solvers but also computationally efficient [116]. Among the existing Riemann solvers, Harten, Lax and Leer [116] proposed the popular solver HLL which is characterized by its simple structure. HLL is an efficient Riemann solver algorithm to compute the flux as in (3.44). Rather than specifying the complete structure of the solution of the Riemann problem, HLL provides an approximation for the interface numerical flux. Let $S_R = \max(\lambda_1, \lambda_2)$ and $S_L = \min(\lambda_1, \lambda_2)$. The HLL Riemann solver at an interface $i + \frac{1}{2}$ is given by

$$U^j_{i+\frac{1}{2}} = \begin{cases} U^j_i & \text{if } S_L \geq 0, \\ U^j_{hull} & \text{if } S_L \leq 0 \leq S_R, \\ U^j_{i+1} & \text{if } S_R \leq 0. \end{cases} \tag{3.45}$$

The HLL solver consists of three states separated by two waves determined by the fastest (S_R) and slowest (S_L) wave speeds. The "hul" state serves as an intermediate state between

the right and left states. The corresponding numerical flux at the interface is given by

$$F^j_{i+\frac{1}{2}} = \begin{cases} F(U^j_i) & \text{if } S_L \geq 0, \\ \frac{S_R F(U^j_i) - S_L F(U^j_{i+1}) + S_L S_R (U^j_{i+1} - U^j_i)}{S_R - S_L} & \text{if } S_L \leq 0 \leq S_R, \\ F(U^j_{i+1}) & \text{if } S_R \leq 0. \end{cases} \tag{3.46}$$

Another popular method to approximate the flux is the Lax-Wendroff. This flux approximation is shown below [116]

$$U^j_{i+\frac{1}{2}} = \frac{\Delta x}{2\Delta t}(F^j_i - F^j_{i+1}) + \frac{1}{2}(U^j_i + U^j_{i+1}), \tag{3.47}$$

$$F^j_{i+\frac{1}{2}} = F(U^j_{i+\frac{1}{2}}) \tag{3.48}$$

In this chapter, we employ both approximation methods to determine the flux at the interface of cells.

3.3.2 Deterministic Inhomogeneous MEM Model

This section considers the deterministic version of (3.1). The source term in the RHS arises naturally in physical applications and represents a sink and/or a source. In equity markets, the source term might represent external (either coming from outside the market or specific to stocks) flux of stocks. In particular, the source term arises when $\rho(x,t)$ and/or $Q(x,t)$ over a price range $[x_1, x_2]$ change due to factors other than the flux (flow of stocks or external forces) through the boundary of the price range.

The system (3.1) can be expressed in a compact form as follows:

$$q(x,t)_t + F(q(x,t))_x = \Psi(q(x,t)), \tag{3.49}$$

where $q(x,t) = (\rho(x,t), Q(x,t))^\top$, $F(q) = (Q(x,t), \frac{(1+\alpha)Q(x,t)^2}{\rho(x,t)})^\top = (q_2, \frac{(1+\alpha)q_2^2}{q_1})^\top$, and

$$\Psi(q) = \begin{bmatrix} a(x,t) + b(x,t)\, q_1 \\ c(x,t) + d(x,t)\, q_2 \end{bmatrix}. \tag{3.50}$$

The source term $\Psi(q)$ ensures that the system of nonlinear partial differential equations approximately describes the equity markets various sources of disturbances. We solve the inhomogeneous form (3.49) by the fractional step method [62], since $\Psi(q)$ depends only on q. This method consists of splitting the problem into a homogeneous conservation law and a simple system of Ordinary Differential Equations (ODEs), and uses standard methods for each. In particular, the homogeneous system (3.11) is first solved by the methods described in Section 3.3.1. Next, the inhomogeneous system reduces to independent systems of ODEs.

$$q_t = \Psi(q). \tag{3.51}$$

The fractional step method based on (3.51) and (3.11), in which $q(x,t)$ is evolved to a value $q(x, t+\Delta)$ over a time step Δt, is summarized as follows

$$\left.\begin{array}{ll} \text{PDEs :} & q_t + F(q(x,t))_x = 0, \\ \text{Initial conditions (ICs) :} & q(x,t), \end{array}\right\} \Rightarrow q^*(x, t+\Delta t), \tag{3.52}$$

$$\left.\begin{array}{ll} \text{ODEs :} & q_t = \Psi(q), \\ \text{ICs :} & q^*(x, t+\Delta t), \end{array}\right\} \Rightarrow q(x, t+\Delta t). \tag{3.53}$$

The main advantage of using the fractional step method is that one can use the best and most efficient existing numerical method for each sub-problem. For example, one can solve the homogeneous sub-problem (3.52) using the Godunov's method. To solve the sub-problem (3.53), there are numerous numerical methods available in ODEs literature. The choice of a solver for the second sub-problem depends on the nature of ODEs. Since the source term $\Psi(q(x,t))$ varies across price-time domain, it might pose challenges such as stiffness. Stiffness exits when the external forces or flux represented by the source term happen on a time scale faster/shorter than the stocks dynamics. The stability and stiffness of the ODEs are determined by its eigenvalues. The eigenvalues of the Jacobian matrix of the first-order system of ODEs (3.53) are given by

$$\frac{\partial \Psi(q(x,t))}{\partial q(x,t)} = \begin{bmatrix} b(x,t) & 0 \\ 0 & d(x,t) \end{bmatrix}, \tag{3.54}$$

where the eigenvalues of (3.54) are $\lambda_1^{ODE} = b(x,t)$ and $\lambda_2^{ODE} = d(x,t)$ [2]. The behavior of (3.53) is determined by the time-varying parameters $b(x,t)$ and $d(x,t)$. The stability condition of the system of ODEs determines the specific ODEs solver, which is specified by the time step Δt_{ODE}. The Courant-Friedrichs-Lewy (CFL) condition (3.43) determines the stability condition for the Godunov's method and specifies the time step Δt associated with (3.52). The system of ODEs is said to be stiff if

$$\max_i \mid \lambda_i^{ODE} \mid \ll \min_j \mid \lambda_j^{ODE} \mid \tag{3.55}$$

For the cases when (3.53) is not stiff, we propose to use a second-order two stage explicit Runge-Kutta method [18, 95] to solve the system of ODEs. This method is a multi-step ODE solver in which midpoints are generated in order to cancel out lower-order error terms. It is stable given a sufficiently small time step. We assume that $\Delta t_{ODE} = \Delta t$. Therefore, in order to evolve $q(x,t)$ at time t to a new value $q(x, t+\Delta t)$ at time $t+\Delta t$, we follow two

[2] The superscript indicates that the eigenvalues are associated with the system of ODEs.

numerical steps based on the Godunov's and Runge-Kutta methods which are implemented sequentially to solve (3.52) and (3.53) as follows:

$$\text{Step } 1 : U_i^* = U_i^j - \frac{\Delta t}{\Delta x}\Big(F\Big(q(U_i^j - U_{i+1}^j)\Big) - F\Big(q(U_{i-1}^j - U_i^j)\Big)\Big), \tag{3.56}$$

$$\text{Step } 2 : k_1 = \Delta t_{ODE}\, \Psi(U_i^*), \tag{3.57}$$

$$k_2 = \Delta t_{ODE}\, \Psi(U_i^* + k_1), \tag{3.58}$$

$$U_i^{j+1} = U_i^* + \frac{1}{2}(k_1 + k_2). \tag{3.59}$$

Stiff ODEs are numerically unstable unless the time step is very small. Implicit methods are commonly used for stiff ODEs so that there are not stability restrictions on the time step. When (3.53) is stiff, we deploy a two-stage Runge-Kutta second-order implicit Method [18, 95] of the form

$$\bar{U}_i = U_i^* + \frac{\Delta t_{ODE}}{2}\Big(\Psi(U_i^*) + \Psi(\bar{U}_i)\Big), \tag{3.60}$$

$$U_i^{j+1} = \frac{1}{3}\Big(4\bar{U}_i - U_i^* + \Delta t_{ODE}\Psi(U_i^{j+1})\Big). \tag{3.61}$$

The previous procedures allow solving both (3.52) and (3.53) in one go. Our previous discussion emphasized the importance of analyzing the stability conditions when using the fractional step method. Careful attention should be paid to the behavior of the system of the ODEs to decide the appropriate ODEs solver. When an implicit method is used, there is no stability condition on the time step Δt_{ODE} and one can assume $\Delta t_{ODE} = \Delta t$. However, when an explicit method is used, a stability analysis must be carried out to determine the time step Δt_{ODE}. If $\Delta t > \Delta t_{ODE}$, the system of ODEs can be solved in d steps, each with size $\Delta t_d = \frac{\Delta t}{d}$, where d is determined such that Δt_d is stable for the ODEs solver. If

$\Delta t \leq \Delta t_{ODE}$, we can simply propagate (3.53) by a time step Δt (i.e., $\Delta t_{ODE} = \Delta t$) .

Given the nature of the source term in the MEM model, one needs to switch between ODEs solvers with caution. A final remark is that the source term is composed of time-varying parameters (and might be stochastic) which vary with market condition. Furthermore, the equity markets data might be noisy and need to be filtered. These challenges motivate us to devise an approach to estimate the source term adaptively. In the next section, we cover the adaptive estimation and filtering processes.

3.4 Stochastic Inhomogeneous MEM Model

This section covers the adaptive estimation of the complete MEM model. Estimation is the process of determining the optimal estimate of the unknown variables (i.e., density, pressure, and parameters) given the observed noisy measurements (stock price data). This process affects the performance of the prediction model since the future evolution of the macroscopic variables depend on the initial condition which needs to be estimated. We propose an algorithm to adaptively predict and update the model states and time-varying parameters. The prediction model is useful in monitoring equity markets. Therefore, accurate and adaptive estimation is required to ensure the reliability of information given to market makers, which helps to achieve faster response to crashes and market anomalies.

In order to study the estimation problem for the MEM model, we review the state-space model. The state-space model representation makes it easier to handle measurement errors in the observed stock price data and provides sequential expressions for the estimation process. In general, the state space model is presented as

$$x_t = g_t(x_{t-1}, w_t), \tag{3.62}$$

$$z_t = h_t(x_t, r_t). \tag{3.63}$$

In equation (3.62), x_t is the state variable of the system at time t, $g_t(\cdot)$ is called the state transition function and w_t is the process noise. Equation (3.63) is called the observation or measurement equation where z_t is the observed variable, $h_t(\cdot)$ is the observation function and r_t is the measurement noise. At every time t, the observation variable z_t is observed and the objective is to determine the best estimation on the state variable x_t.

Let the functions g and h in (3.62) and (3.63), respectively, be linear. For instance, consider the state-space model in (3.64).

$$\begin{aligned} x_t &= G_t x_{x-1} + v_t, \\ z_t &= H_t x_t + r_t, \end{aligned} \tag{3.64}$$

where x_t is the state variable at time t, z_t is the observed variable with measurement error $r_t \sim \mathcal{N}(0, V_t)$. The state transition and observation models are denoted by G_t and H_t, respectively. The process noise is $w_t \sim \mathcal{N}(0, R_t)$.

The optimal estimation of the linear state space model (3.64) can be achieved by Kalman Filtering. Kalman filtering (KF) [68] is a recursive method that determines the optimal estimate of x_t given the measurement z_t. The KF algorithm works in two steps: prediction and update. In the prediction step, the KF *predicts* the current state variable $\hat{x}_{t|t-1}$ and error covariance $P_{t|t-1}$ given the previous posterior estimates of the state variable $\hat{x}_{t-1|t-1}$ and error covariance $P_{t-1|t-1}$. Once the outcome of the noisy measurement y_t is observed at time t, the KF *updates* the estimates to produce the posterior estimates of the state variable $\hat{x}_{t|t}$ and error covariance $P_{t|t}$. The complete KL algorithm is outlined below.

Prediction Step:

$$\hat{x}_{t|t-1} = G_t \hat{x}_{t-1|t-1}, \tag{3.65}$$

$$P_{t|t-1} = G_t P_{t-1|t-1} G_t^\top + V_t. \tag{3.66}$$

Update Step:

$$y_t = z_t - H_t \hat{x}_{t|t-1}, \tag{3.67}$$

$$K_t = P_{t|t-1} H_t^\top (H_t P_{t|t-1} H_t^\top + R_t)^{-1}, \tag{3.68}$$

$$\hat{x}_{t|t} = \hat{x}_{t|t-1} + K_t y_t, \tag{3.69}$$

$$P_{t|t} = (I - K_t R_t) P_{t|t-1}, \tag{3.70}$$

where K_t is the optimal gain and y_t is the measurement residual.

3.4.1 State-Space Representation of the MEM Model

A recursive estimation method is needed to update current equity markets state and estimation of parameters and make one-step ahead prediction in real-time when new observations are collected. To begin, we propose a discretized version of the MEM in a state-space model. Suppose that the price domain starting at 0 and ending at x_N is divided into N segments of length Δx. Similarly, the time domain starting at time 0 and ending at time t_M is divided into M segments of length Δt . This discretization process is similar to the domain construction discussed in Section 3.3. The macroscopic variables density $\rho(x,t)$ and average velocity $v(x,t)$ are defined at discrete time intervals Δt and price intervals Δx. We assume that Δx is one dollar and Δt is one minute. These choices of price and time step lengths satisfy the CFL condition (3.43) in Godunov's scheme.

The estimation process of the MEM model includes estimating the stocks density $\rho(x_i, t_j)$, flux $Q(x_i, t_j)$ and model parameters represented by the source term vector $\Psi(q(x_i, t_j))$ at time $t_j \geq t$ for $i = 1, 2, \cdots, N$, $j = 1, 2, \cdots, M$, given the observations up to time t. Let ρ_i^j denote the density in price i at time $j\Delta t$ and

$$\boldsymbol{\rho}^j = [\rho_1^j, \rho_2^j, \cdots, \rho_N^j]^\top. \tag{3.71}$$

The flux $\boldsymbol{Q}^j$ is defined similarly as the flux vector at time $j\Delta t$. Let $\boldsymbol{\theta}^j$, $\boldsymbol{\pi}^j$, and $\boldsymbol{\phi}^j$ denote the parameters vectors at time $j\Delta t$, as shown in equations (3.72) [3]

$$\begin{aligned}
\boldsymbol{\theta}^j &= [a_1^j, b_1^j, a_2^j, b_1^j, \cdots, a_N^j, a_N^j, \sigma_{1,N}^j]^\top, \\
\boldsymbol{\pi}^j &= [c_1^j, d_1^j, c_2^j, d_1^j, \cdots, c_N^j, d_N^j]^\top, \\
\boldsymbol{\phi}^j &= [\sigma_{1,1}^j, \sigma_{1,2}^j, \cdots, \sigma_{1,N}^j, \sigma_{2,1}^j, \sigma_{2,2}^j, \cdots, \sigma_{2,N}^j]^\top.
\end{aligned} \tag{3.72}$$

Given $\boldsymbol{\rho}^{j-1}$, $\boldsymbol{Q}^{j-1}$, $\boldsymbol{\theta}^j$, and $\boldsymbol{\pi}^j$, the numerical procedure in (3.56) and (3.57) determines the one-ahead step prediction of $\boldsymbol{\rho}^j$ and $\boldsymbol{Q}^j$. We rewrite the numerical procedure to solve the deterministic inhomogeneous MEM as (3.73), where $\boldsymbol{q}^j = (\boldsymbol{\rho}^j, \boldsymbol{Q}^j)^\top$, $\boldsymbol{\eta}^j = (\boldsymbol{\theta}^j, \boldsymbol{\pi}^j, \boldsymbol{\phi}^j)^\top$, and $\boldsymbol{v}$ is the noise vector. The noise term represented by the Brownian sheet in each equation of (3.1) consists of N i.i.d normal random variables with mean 0 and standard deviation $1/\sqrt{2}$. Therefore, $\boldsymbol{v}$ is a $2N$-dimensional vector.

$$\boldsymbol{q}^j = g(\boldsymbol{q}^{j-1}, \boldsymbol{\eta}^j, \boldsymbol{v}), \tag{3.73}$$

In the functional form (3.73), g is the state transition function. This function does not have a closed form due to the complexity of the numerical method used in this chapter. However, g can be obtained by numerical procedure and this is sufficient for the estimation process to work. Since the evolution of the state variable $\mathbf{q}^j$ depends on the parameters as shown in (3.73), the estimation of the parameters is important. In Chapter 2, the parameters are estimated in an off-line fashion which incurs high computational costs. The underlying assumption there is that the parameters are constant over a number of minutes. This assumption might be valid during regular trading days and for a very short period of time. However, during flash crashes, equity markets undergo dramatic shifts in stocks dynamics. This motivates the need for an adaptive algorithm to capture, in a short period of time, the changes in market conditions as well as determine their influence on the model's parameters.

[3] The volatility terms σ_1 and σ_2 have two subscripts in equations (3.72). The first subscript indicates the volatility type (the first type belongs to the first equation in (3.1) while the second type belongs to the section equation), whereas the second subscript indicates the price location i.

In order to estimate the parameters, we assume that they evolve according to a random walk model as shown below

$$\boldsymbol{\eta}^j = \boldsymbol{\eta}^{j-1} + \boldsymbol{w}^j, \tag{3.74}$$

where $\boldsymbol{w}^j$ is a normally distributed random vector.

To have a complete state space model as in equations (3.62) and (3.63), we further specify the measurement equation. Let z_i^j denote the observed density and flux in price i at time $j\Delta t$, and $\boldsymbol{z}^j$ denote the observation variable vector at $j\Delta t$ as shown in equation (3.75).

$$\boldsymbol{z}^j = [z_1^j, z_2^j, \cdots, z_N^j]^\top \tag{3.75}$$

Since the measurements are usually noisy, the measurement equation for the observed variables is assumed to be given by equation (3.76).

$$\boldsymbol{z}^j = h(\boldsymbol{q}^j, \boldsymbol{r}), \tag{3.76}$$

where $\boldsymbol{r}$ is a $2N$-dimensional measurement noise vector, and h is a measurement function. The complete state space model for the macroscopic equity market model is formulated as follows

$$\begin{aligned} \boldsymbol{\eta}^j &= \boldsymbol{\eta}^{j-1} + \boldsymbol{w}^j, \\ \boldsymbol{q}^j &= g(\boldsymbol{q}^{j-1}, \boldsymbol{\eta}^j, \boldsymbol{v}), \\ \boldsymbol{z}^j &= h(\boldsymbol{q}^j, \boldsymbol{r}). \end{aligned} \tag{3.77}$$

Let $\hat{\boldsymbol{q}}(j \mid j-1)$ denote the prior estimation of the state variable $\boldsymbol{q}^j$ at time j given the observations up to and including time $j-1$ (i.e., prior mean estimate), $P_{\boldsymbol{q}}(j \mid j-1)$ denotes the prior error covariance matrix of the state estimation up to and including time $j-1$ (i.e.,

prior covariance estimate), $\hat{\boldsymbol{q}}(j \mid j)$ denotes the posterior estimation of the state variable vector $\boldsymbol{q}^j$ at time j given the observations up to and including time j (i.e., posterior mean estimate), and $P_{\boldsymbol{q}}(j \mid j)$ denotes the posterior error covariance matrix of the state estimation up to and including time j (i.e., posterior covariance estimate). Similar notions are used for $\boldsymbol{\eta}^j$ and $\boldsymbol{z}^j$ to denote the mean and covariance estimates. We also assume that the noise vectors are described as follows: $\boldsymbol{w}^j \sim \mathcal{N}(0, P_{\boldsymbol{w}}^j)$, $\boldsymbol{v} \sim \mathcal{N}(0, P_{\boldsymbol{v}})$, and $\boldsymbol{r} \sim \mathcal{N}(0, P_{\boldsymbol{r}})$, where $P_{\boldsymbol{v}} = 0.5 I_{2N}$ and $I_{2N} \in \mathbb{R}^{2N \times 2N}$ is an identity matrix. The error covariance matrices measure the accuracy of the state estimation.

Throughout this chapter, we make the following assumptions. 1) The measurement function h is linear in the observed state variable $\boldsymbol{q}$. 2) The noise in both the transition g and the measurement equations h is additive. In order to obtain the optimal estimate $\hat{\boldsymbol{q}}(j \mid j)$, we implement a recursive estimation method known as filtration. Inspired by KF, the estimation process (formally known as filtering) consists of two steps: first, a time update or prediction step is performed in which the previous observations and the state transition function are used to estimate the state variable $\hat{\boldsymbol{q}}(j \mid j-1)$. Second, a measurement or filtering update step in which the current measurements or observations are utilized to refine (filter) the estimated state variable $\hat{\boldsymbol{q}}(j \mid j)$. A complete treatment about the estimation problem can be found in standard books on optimal estimation (see e.g. [49]).

In the next two subsections, we cover different filtering techniques that we implement to solve the estimation problem (3.77). In the first subsection, we introduce two methods for estimating the state variable given that the parameters are constant. In the second subsection, we approach the problem differently by combining both parameters and state estimations in a dual fashion. This dual estimation technique is an extension of the filtering methods for state estimation but the estimated variable is not only the state of the model but, also the parameters. Other methods exist in the optimal estimation literature for parameters estimation. These methods are closely related to machine learning algorithms where a training set (input and output) is available to solve for the parameters such that the

expected square error for the machine (difference between the true and expected output) is minimized [121]. The dual estimation approach is preferred in our case since such data is, in fact, not observed.

3.4.1.1 Estimation of q

Though KF obtains the optimal estimation of the linear state-space model, the complexity of g in (3.77) is nonlinear. The Extended Kalman Filter (EKF) is considered the standard estimation technique for nonlinear models in which the nonlinear model is linearized so that the classical linear KF can be implemented. In this filter, the state distribution is approximated by a Gaussian Random Variable (GRV) and then is propagated by the first-order linearization of the nonlinear model. However, the EKF is difficult to implement and tune [67], mainly due to two drawbacks: the linearisation might produce instability in the filters, and the calculation of the derivatives (while approximating the nonlinear model by a linear one) is nontrivial in most practical applications.

The authors of [67] proposed a superior algorithm, known as the Unscented Kalman Filter (UKF), that overcomes the drawbacks of EKF and more importantly, calculating the derivatives is not required. The central idea of the UKF is based on the unscented transformation. In UKF, the state variable distribution is also approximated by a number of sample points (called sigma points hereafter). The sigma points are chosen carefully to capture the true mean and covariance of the state distribution. These sigma points are propagated through the true nonlinear transformation. The propagated sigma points are used to calculate the posterior mean and covariance. The UKF is reported to approximate the posterior mean and covariance of any nonlinear models up to the third order (compared to a Taylor series expansion) [67] for Gaussian random variables.

Before we present the UKF algorithm, we briefly describe the Unscented Transformation (UT) method, which is its central idea. The UT is a method for finding statistics such as mean and covariance of a random variable which is subjected to a nonlinear transformation.

For instance, consider a random variable $\boldsymbol{y} = f(\boldsymbol{x})$ where $f(\cdot)$ is a nonlinear function and $\boldsymbol{x} \in \mathbb{R}^L$ is a random variable with mean $\bar{\boldsymbol{x}}$ and covariance $P_{\boldsymbol{x}}$. In order to approximate the mean and covariance of $\boldsymbol{y}$, a set of $2L+1$ sigma vectors $\mathcal{S}_i$ $(i = 0, 1, \cdots, 2L)$ are generated as follows:

$$\mathcal{S}_0 = \bar{\boldsymbol{x}}, \tag{3.78}$$

$$\mathcal{S}_i = \bar{\boldsymbol{x}} + \left(\sqrt{(L+\lambda)P_{\boldsymbol{x}}}\right)_i, \quad i = 1, \cdots, L \tag{3.79}$$

$$\mathcal{S}_{i+L} = \bar{\boldsymbol{x}} - \left(\sqrt{(L+\lambda)P_{\boldsymbol{x}}}\right)_{i+L}, \quad i = 1, \cdots, L \tag{3.80}$$

$$\tag{3.81}$$

where L is the dimension of the random variable $\boldsymbol{x}$ and $\lambda = \alpha^2(L+\kappa) - L$ is a scaling factor. The parameters α and κ control the spread of the sigma points, which are normally set to 0.001 and 0 respectively. The term $\left(\sqrt{(L+\lambda)P_{\boldsymbol{x}}}\right)_i$ is the ith column of the square root of the matrix $\left((L+\lambda)P_{\boldsymbol{x}}\right)$. Each sigma vector $\mathcal{S}_i$ is propagated through the nonlinear function $g(\cdot)$ to obtain posterior sigma points $\mathcal{Y}_i = g(\mathcal{S}_i), i = 0, \cdots, 2L$. The statistics of the random variable y (mean and covariance) are approximated by the sample weighted mean and covariance. The weights W^i (associated with $\mathcal{S}^i$ for $i = 0, 1, \cdots, 2L$) are given by:

$$W_0 = \frac{\lambda}{L+\lambda}, \tag{3.82}$$

$$W_i = \frac{1}{2(L+\lambda)} \quad i = i = 1, \cdots, 2L. \tag{3.83}$$

$$\tag{3.84}$$

Turning to our estimation problem, the recursive estimation by UKL algorithm (prediction and updating steps) of the random state vector $\boldsymbol{q} \in \mathbb{R}^{2N}$ in (3.77) is shown below. It should be noted that the parameters in the following algorithm and subsequent one are assumed to

be known.

Algorithm 1

Initialize at time $j = 0$

$\hat{\boldsymbol{q}}(0 \mid 0) = \mathbb{E}[\boldsymbol{q}_0]$, $P_{\boldsymbol{q}}(0 \mid 0) = \mathbb{E}\big[\big(\boldsymbol{q} - \hat{\boldsymbol{q}}(0 \mid 0)\big)\big(\boldsymbol{q} - \hat{\boldsymbol{q}}(0 \mid 0)\big)^\top\big]$

$j \leftarrow j + 1$

while $j < M$ **do**

  Compute sigma vectors: $\mathcal{S}(j) = \Big[\hat{\boldsymbol{q}}(j \mid j),\ \ \hat{\boldsymbol{q}}(j \mid j) + \big(\sqrt{(N+\lambda)P_{\boldsymbol{q}}(j \mid j)}\big),\ \ \hat{\boldsymbol{q}}(j \mid j) - \big(\sqrt{(N+\lambda)P_{\boldsymbol{q}}(j \mid j)}\big)\Big]$

  Time update

    $\mathcal{S}(j+1 \mid j) = g(\mathcal{S}(j), \boldsymbol{\theta}, \boldsymbol{\pi})$

    $\hat{\boldsymbol{q}}(j+1 \mid j) = \sum_{i=0}^{4N} W_i \ \mathcal{S}_i(j+1 \mid j)$

    $P_{\boldsymbol{q}}(j+1 \mid j) = \sum_{i=0}^{4N} W_i \big[\mathcal{S}_i(j) - \hat{\boldsymbol{q}}(j+1 \mid j)\big]\big[\mathcal{S}_i(j) - \hat{\boldsymbol{q}}(j+1 \mid j)\big]^\top + \boldsymbol{\phi}^{2^\top} P_{\boldsymbol{v}}$

    $\mathcal{Z}(j+1 \mid j) = h(\mathcal{S}(j))$

    $\hat{\boldsymbol{z}}(j+1 \mid j) = \sum_{i=0}^{4N} W_i \mathcal{Z}_i(j+1 \mid j)$

  Measurement update

    $P_{\boldsymbol{z}}(j+1 \mid j) = \sum_{i=0}^{4N} W_i \big[\mathcal{Z}_i(j+1 \mid j) - \hat{\boldsymbol{z}}(j+1 \mid j)\big]\big[\mathcal{Z}_i(j+1 \mid j) - \hat{\boldsymbol{z}}(j+1 \mid j)\big]^\top + P_{\boldsymbol{r}}$

    $P_{\boldsymbol{qz}}(j+1 \mid j) = \sum_{i=0}^{2N} W_i \big[\mathcal{S}_i(j) - \hat{\boldsymbol{q}}(j+1 \mid j)\big]\big[\mathcal{Z}_i(j+1 \mid j) - \hat{\boldsymbol{z}}(j+1 \mid j)\big]^\top$

    $K(j+1) = P_{\boldsymbol{qz}}(j+1 \mid j) P_{\boldsymbol{z}}^{-1}(j+1 \mid j)$

    $\hat{\boldsymbol{q}}(j+1 \mid j+1) = \hat{\boldsymbol{q}}(j+1 \mid j) + K(j+1)(\boldsymbol{z}(j+1) - \hat{\boldsymbol{z}}(j+1 \mid j))$

    $P_{\boldsymbol{q}}(j+1 \mid j+1) = P_{\boldsymbol{q}}(j+1 \mid j) - K(j+1) P_{\boldsymbol{z}}(j+1 \mid j) K^\top(j+1)$

  $j \leftarrow j + 1$

end while

We propose to implement another filter that has been recently introduced to the statistical community. The Ensemble Kalman Filter (EnKF) was first introduced in [42] as an alternative filtering approach to complex problems with strong nonlinearity. This method is popular due to its flexibility and efficiency [70]. The EnKF has several extensions that have been introduced in geophysics literature (i.e., [115] and [11]) where real-life problems exhibit high complexity and high dimensionality. In many of the geophysics applications, this method is reported to yield very competitive results compared to the EKF and other filtering methods [16].

The EnKF is a sequential filtering method that is suitable for problems involving discretization of a system of partial differential equations. This method is considered a Monte-Carlo implementation to the Bayesian problem in which the prior estimation with new data is used to obtain the posterior estimation. The EnKF works as follows: a number of sample points called ensembles are generated from the prior estimation distribution. Each ensemble point is updated using a linear rule and Kalman gain, similar to the original KF update equation. The posterior estimation is the mean of the updated ensembles while their spread is the posterior covariance.

There are mainly two approaches for updating the ensembles in EnFK, stochastic and deterministic updates. In stochastic update, the updated ensembles are adjusted by a stochastic perturbation. While in deterministic update, ensembles are updated by a non-random transformation of the forecast ensembles (see [70], for more details). In this chapter, we only apply the stochastic update in the EKF since it produces more robust estimates [79, 77]. The stochastic EnKF algorithm for estimating $\boldsymbol{q}$ is shown below.

Algorithm 2

Initialize at time $j = 0$ with the ensembles $\hat{\boldsymbol{q}}^1(0 \mid 0), \cdots, \hat{\boldsymbol{q}}^L(0 \mid 0)$
$j \leftarrow j + 1$
while $j < M$ **do**
    for $l = 1, \cdots, L$ **do**
        Time update: Draw sample $\boldsymbol{v}^l \sim \mathcal{N}(0, P_{\boldsymbol{v}})$
            $\hat{\boldsymbol{q}}^l(j+1 \mid j) = g\big(\hat{\boldsymbol{q}}^l(j \mid j), \boldsymbol{\theta}, \boldsymbol{\pi}\big) + \boldsymbol{\phi}^\top \boldsymbol{v}^l(j)$
    end for
    for $l = 1, \cdots, L$ **do**
        Measurement update: Draw sample $\boldsymbol{r}^l \sim \mathcal{N}(0, P_{\boldsymbol{r}})$
        $\hat{\boldsymbol{q}}^l(j+1 \mid j+1) = \hat{\boldsymbol{q}}^l(j+1 \mid j) + K(j+1)\Big(\boldsymbol{z}(j+1) - h\big(\hat{\boldsymbol{q}}^l(j+1 \mid j), \boldsymbol{r}^l\big)\Big)$
    end for
    $j \leftarrow j + 1$
end while

3.4.1.2 State and Parameters (Dual) Estimations

Dual estimation problems consist of estimating the state variable $\boldsymbol{q}^j$ and the model parameter $\boldsymbol{\eta}^j$ simultaneously from the noisy observed data $\boldsymbol{z}^j$. The dual problem has been approached from different angels and many methods exist for solving it [57]. We propose a general paradigm applicable to both filtering methods presented in the previous section. Assume that we have a separate state-space model for the model's parameters as shown below

$$\begin{aligned} \boldsymbol{\eta}^j &= \boldsymbol{\eta}^{j-1} + \boldsymbol{w}^j, \\ \boldsymbol{z}^j &= h(\boldsymbol{q}^{j-1}, \boldsymbol{r}), \end{aligned} \tag{3.85}$$

where the innovations $\boldsymbol{w}^j$ and $\boldsymbol{r}$ are independent normally distributed random variables with means zero and covariances $P_{\boldsymbol{w}}^j$ and $P_{\boldsymbol{r}}$, respectively. In the dual estimation, two filtering methods, say for instance UKF, run simultaneously for the state variable and parameters. At every time step j, the current estimate for the state variable is used to estimate the parameters, while the current estimate for the parameters are used to estimate the state variable. Both estimation problems are solved by the same filtering method [121].

3.5 Results

This section analyzes the application of the proposed methods on normal trading days and abnormal ones. On normal trading days, the equity markets are stable. Abnormal trading days can be associated with sharp and/or sudden fluctuations which are often attributed to unknown or external factors/forces. We implement the proposed methods to determine the time evolution of density and flux of stocks. In addition to analyzing the macroscopic variables, the model's parameters are evaluated as well. To this end, this section integrates the PDEs solvers and filtering techniques proposed in the previous section to dynamically solve and determine the evolution of density ρ, flux Q and forcing terms (parameters).

3.5.1 Data

The data used in this analysis is composed of three equity markets: NASDAQ, NYSE, AMEX. The raw data comprises minute resolution stock price information over a trading period of 9 hours per day. Close to 4000 stocks were collected. The space and time domains were discretized according to the CFL condition. That is, space (price) is discretized into N fixed intervals of size Δx and time is discretized into M fixed periods of length Δt. The CFL stability condition requires an upper bound on the ratio $\Delta x/\Delta t$, which relates to the maximum information speed λ. The information speed is a function of stock velocity, which is roughly bounded by one. This bound was validated by separate data. The average stocks velocity $v(x_i, t_j)$ and $\rho(x_i, t_j)$ are computed at price x_i and time t_j where $i = 0, \cdots, N$ and $j = 0, \cdots, M$.

3.5.2 Implementation

In the previous sections, we proposed several methods to solve the MEM model sequentially. The evaluation of density and velocity are determined by using numerical methods to solve the deterministic inhomogeneous MEM model. The stochastic inhomogeneous MEM is solved using stochastic filtering methods. We proposed two approaches for both problems: the deterministic and stochastic models. The proposed methods are HLL and Lax-Waderoff for solving the deterministic problem, and UKL and EnKL to tackle the stochastic version of the model. In this section, we illustrate the implantation procedure of employing HLL and UKL to solve the complete MEM model. The procedure is similar for the case when the other numerical solver (i.e., Lax-Waderoff) and stochastic filter method (i.e., EnKL) are applied.

The estimation problem of the state-space model in (3.77) includes estimating, sequentially, the expected value and covariance of $\boldsymbol{q}$ and $\boldsymbol{\eta}$. Initially, the data in terms of the temporal axis, was divided into three sets: training, testing, and validation sets. The tem-

poral split of data at time j is shown in Figure 3.4.

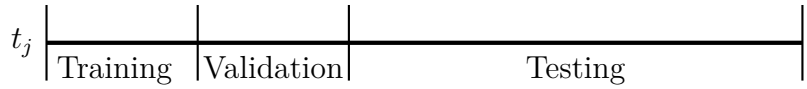


Figure 3.4: The Temporal Split of Data at Time j

The first part is used to compute the initial values of the states and parameters. The initial values can be computed as line 2 of Algorithm 1. We should mention that the dual approach is taken in this chapter. That is to say, Algorithm 1 is used for both the states and parameters estimation. When the initial values for $\boldsymbol{q}$ and $\boldsymbol{\eta}$ are computed, starting from time $j = 1$ in the testing set, the time update is performed as in lines 7-11 in the Algorithm. The parameters time update was done first and these values are used in the update of the state, specifically in line 7. In the time update, we mentioned that the function g is complex but we suggested to use the numerical procedure (3.56) and (3.57) to determine the predictions. The predicted values of the parameters are used in (3.57). From the numerical procedure, we see that the flux function approximation is needed. We proposed two methods and we report the results of using both approximations in the next section. When the density and flux are observed, their values are updated according to lines 13–17. A similar updating scheme is used to update the parameters from the observed density and flux. The update step in line 16 should be based on the observed state variables since the parameters are not observed. This remark can be seen from the state-space model of the parameters in (3.85). The previous implementation is repeated sequentially, in a discrete time step of size one. Keeping the previous steps in mind, the validation dataset is used to validate the tuning parameters, in the state-space model, which are the states and parameters measurement errors.

The validation over multiple trading days resulted in several observations. The volatility of predicted density increases as the density measurement error increases. We observed that the length of training period should be about 25 minutes. We also noticed that density mea-

surement error should not be too high as that would result in a very wide confidence bound for the forecasts. The parameters measurement error with value equal to zero generated mixed results. When that parameters measurement error was between 0.001 and 0.5, the minimum average Root Mean Square Error (RMSE) across all the price ranges was obtained. It should be noted that, in our final implementation, we used the following setting: the flux and density measurement errors were 0.1 and 0.8, respectively.

3.5.3 Case Study 1: The Flash Crash

We first implement the integrated algorithm on data collected from the flash crash day, May 6th 2010. This day is considered by many researchers as unique in that we first observed consequences of the complexity of the algorithmic trading system. On that day, several markets experienced a sudden sharp decline followed by a correction in which losses were recovered by the end of the trading day. This phenomenon has been the focuse of many important studies [38]. Some explanations for the sharp decline and recovery include: the large purchase of put options, large sell of E-Mini contracts, and quote stuffing . However, the actual causes of the flash crash are a matter of debate. Our study does not attempt to explain the causes of this phenomenon, but provides an early detection method.

To begin, we look more closely at the flash crash data by plotting the velocity and density to see if there is anything noticeable. Figure 2 shows the density of stocks (i.e., the number of stocks per price) over a period of 391 minutes, covering the trading period from 9:30 a.m. and 4:00 p.m. In the color map to the right of the figure, the greater the density, the darker the color. As can be seen, the stocks are concentrated on the first 40 price ranges and then the density reduces gradually. This observation is highly reasonable as we expect that the number of stocks to gradually decrease as the price increases. The important question now is, which price ranges will be the focus of our study? Following insights from fluid dynamics, the fluid description is more appropriate when the number of particles is large enough. Similarly, in order to maintain validity in equity markets, we consider only the first

25 price ranges. In these ranges, the density appears to be high and stocks are more actively traded (note how the density changes over time).

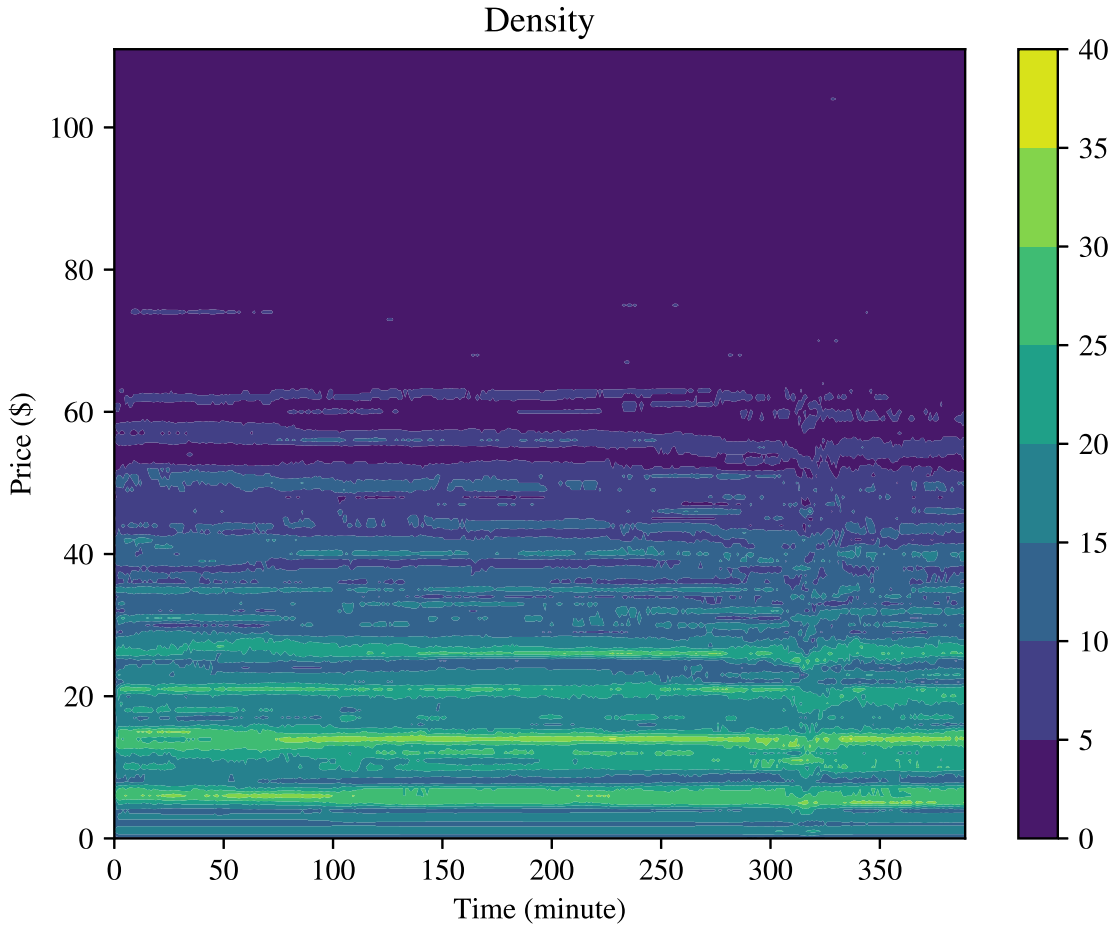


Figure 3.5: Density on the Flash Crash Day

Knowing that the flash crash started at 02:32 p.m. (302 minutes in the x-axis), it can be seen that before that time, nothing unusual is observed. Around the flash crash time, a sharp decline is very noticeable. The plunge persists over most of the price ranges. In other words, the disturbance propagates in a certain direction, affecting all stocks. This behavior is well captured by our model through what we defined as shocks and rarefaction waves.

Turning to the velocity of stocks, the velocity figures for various price ranges are illustrated in Figure 3. The velocity appears to offer more information and sensitivity around the flash crash time. Although this might not be obvious from the figure, the disturbance is also persistent across various price ranges. More specifically, shock or rarefaction waves occurred in the velocity of stocks. From our examination of the density and velocity, we see that the macroscopic variables are best suited for signaling abnormality. Given this information, it is still interesting to see the performance of the proposed algorithms on predicting these variables.

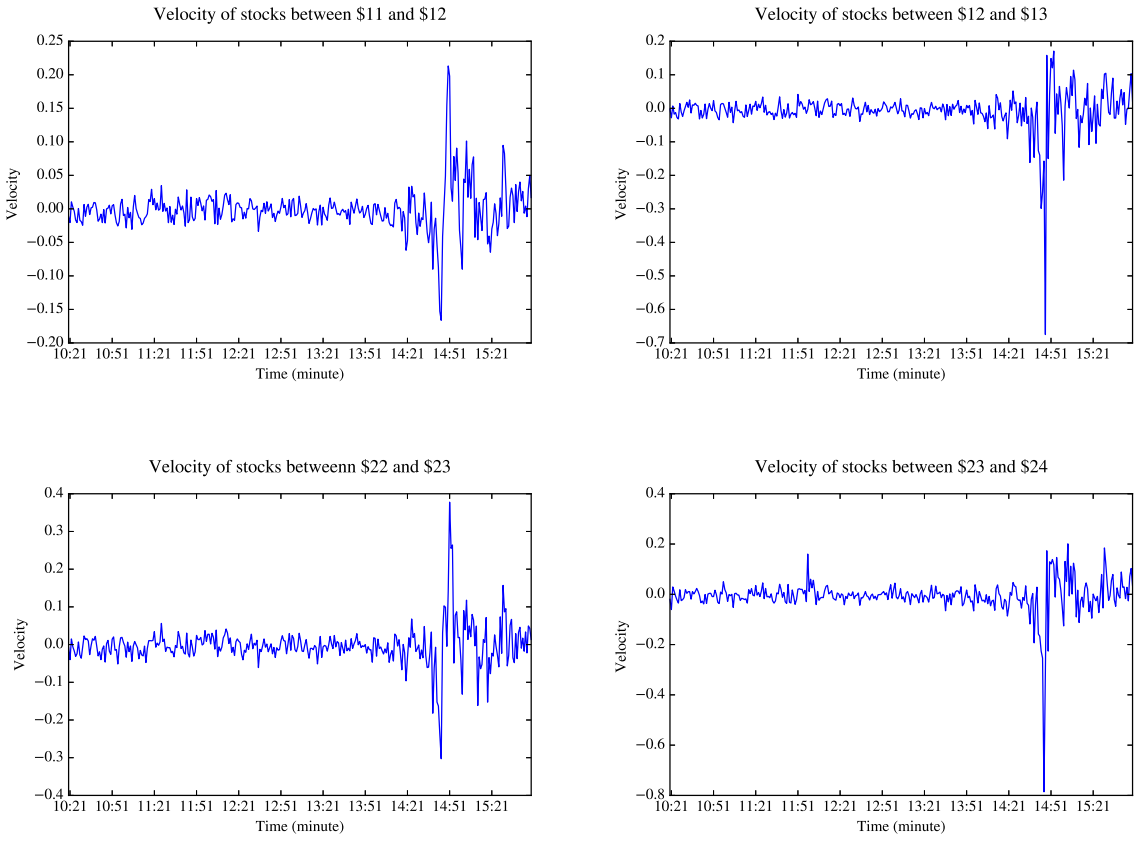


Figure 3.6: Velocity Plots

We present and discuss the prediction results of implementing the proposed algorithms

on only a few price points i. In particular, we select four price ranges to evaluate whether the uncertainty exhibited by the data is well captured by the model. Also, we analyze the prediction model using two performance measures: Root Mean Square Error (RMSE) and Mean Square Error (MSE). The table below presents the average RMSE and MSE over the first 25 price points. Table 3.1 shows that the Lax-Wendroff outperforms the other methods.

Model		RMSE	MSE
PDEs Solver	Stochastic Filtering Technique		
HLL	Dual UKL	1.94	1.35
HLL	Dual EnKL	1.91	1.31
Lax-Wendroff	Dual UKL	1.80	1.31
Lax-Wendroff	Dual EnKL	1.92	1.31

Table 3.1: Average Performance of the Proposed Algorithms in Predicting the Density of Stocks

Model		RMSE	MSE
PDEs Solver	Stochastic Filtering Technique		
HLL	Dual UKL	1.02	0.55
HLL	Dual EnKL	1.38	0.59
Lax-Wendroff	Dual UKL	0.96	0.53
Lax-Wendroff	Dual EnKL	1.35	0.59

Table 3.2: Average Performance of the Proposed Algorithms in Predicting the Flux of Stocks

The figures below demonstrate the predicted flux and density for the period between 12:49 p.m. and 03:49 p.m. Figures 3.8–3.7 show that the model is able to predict the direction of the crash and shows some signals prior to the crash. As shown in Figures 3.8–3.7, most of the observations are within the uncertainty bounds even during high volatility periods.

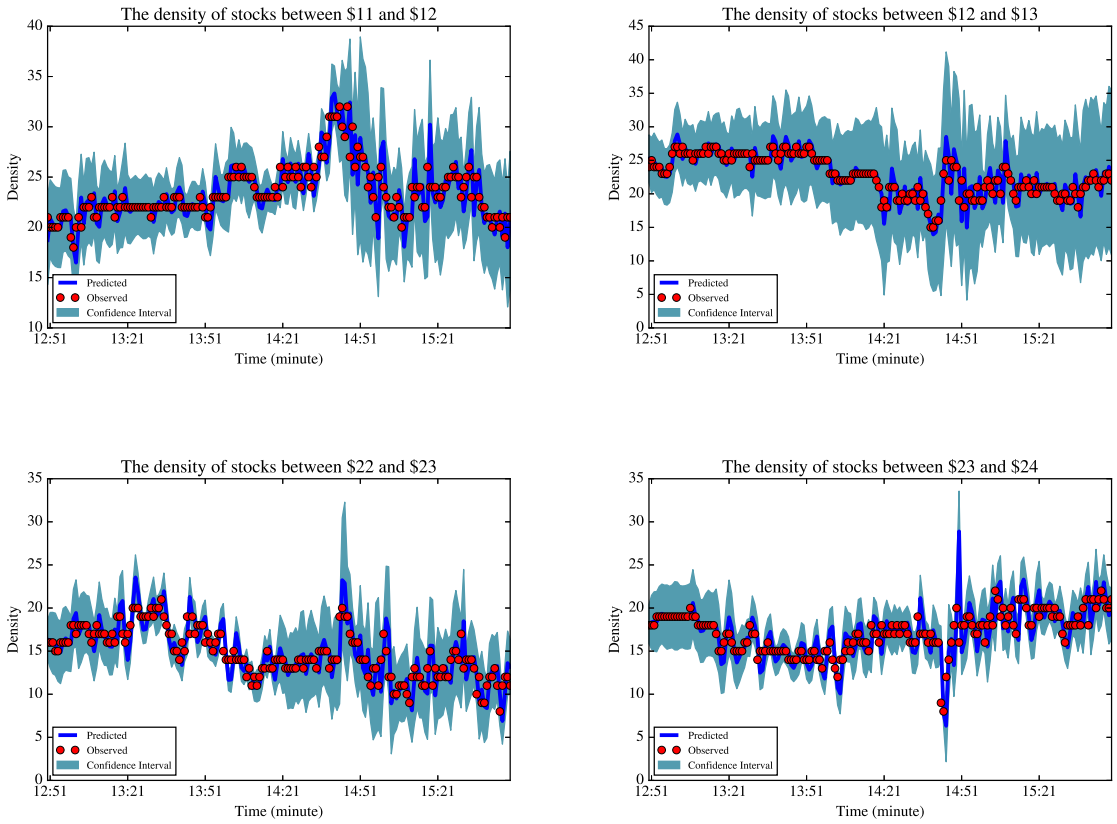


Figure 3.7: One-Step Ahead Prediction of Density

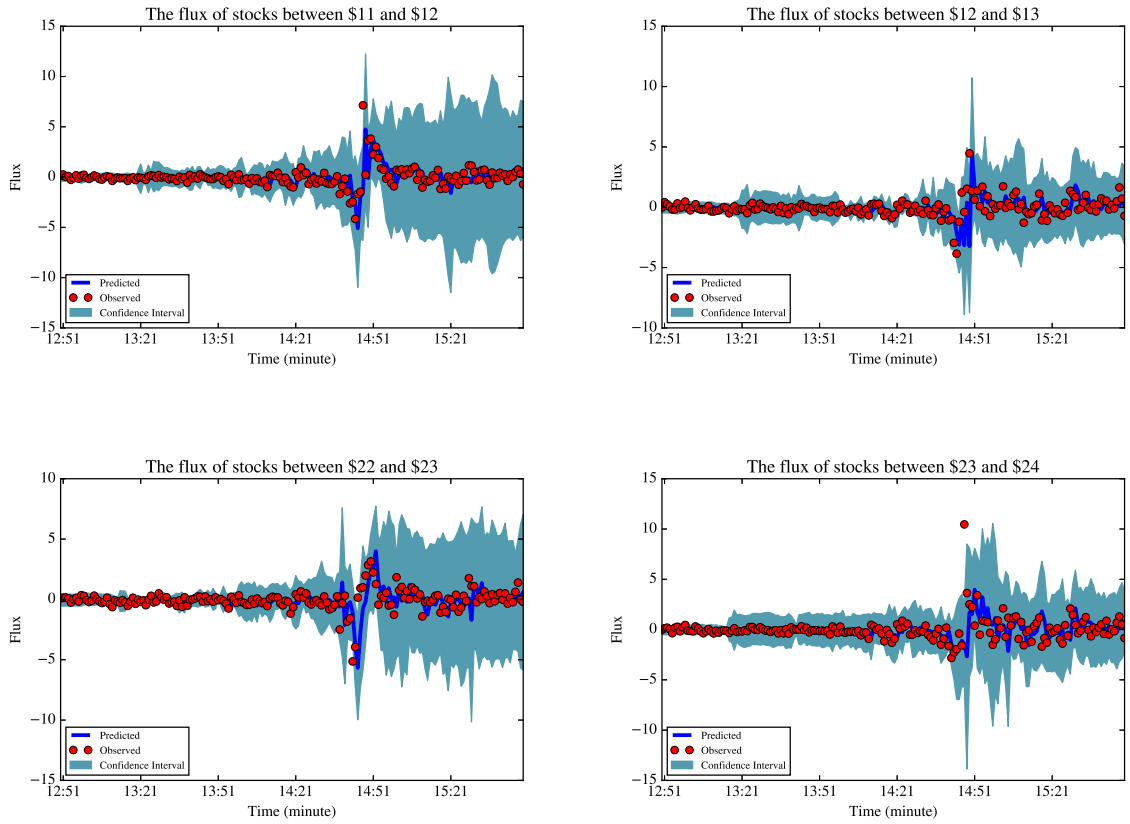


Figure 3.8: One-Step Ahead Prediction of Flux

Even though the previous results are informative, we present the prediction results for what we call the "pressure". The pressure is defined as αQv. We believe that this measure is more sensitive to crashes and abnormalities in the market. As can be seen in the figures below, the predicted pressure shows high sensitivity around the flash crash time and afterwards.

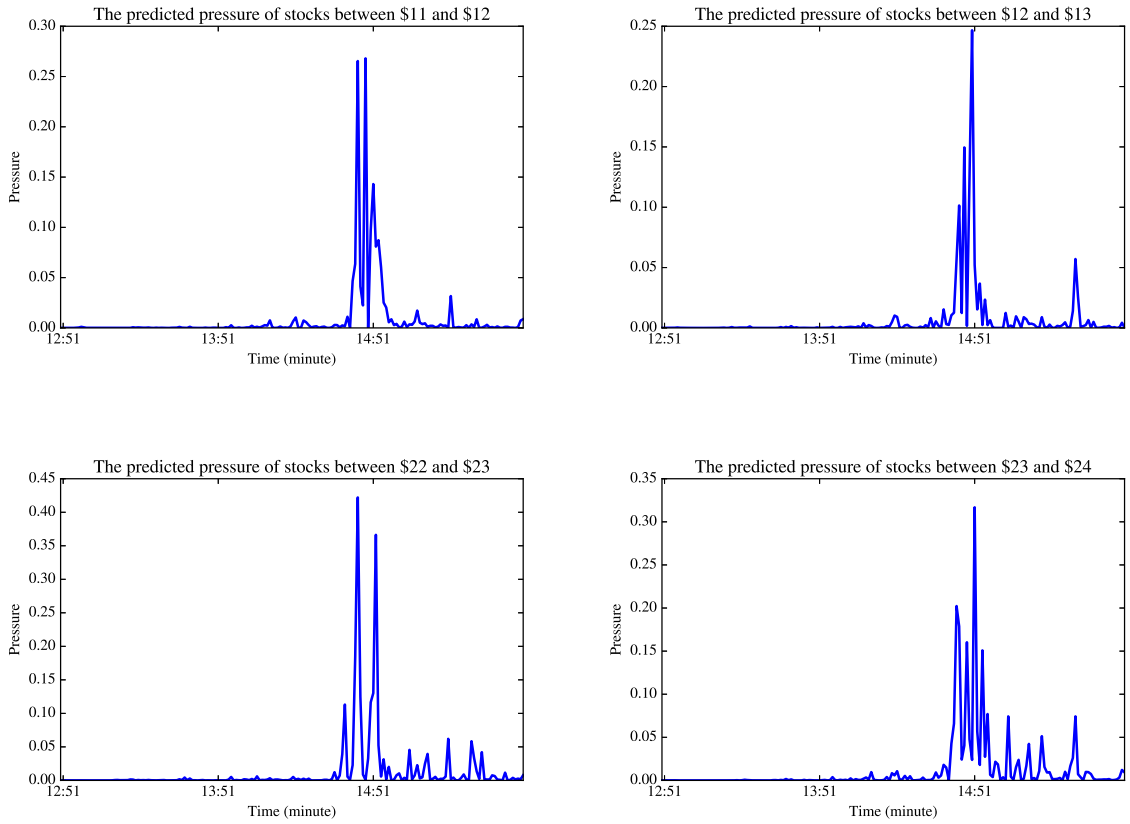


Figure 3.9: One-Step Ahead Prediction of Pressure

3.5.4 Case Study 2: Facebook's Initial Public Offering

In this section, we consider another trading day. We implemented our proposed methodology on data obtained from May 18th, 2012. On that day, Facebook had its initial public offering on the NASDAQ stock market to raise $16 billion [100]. It was considered one of the biggest offerings ever in the U.S. This massive offering showed to be a problem for NASDAQ. The market experienced technical difficulties, delaying the trading of the stock for about 30 minute which left traders stranded. These difficulties were reported to be caused by glitches in the market software. The glitches created turmoil for a number of market makers for

Facebook's stock, resulting in a total lose of $115 million. Similar to our analysis on the flash crash day, we show the velocity plot for several price ranges.

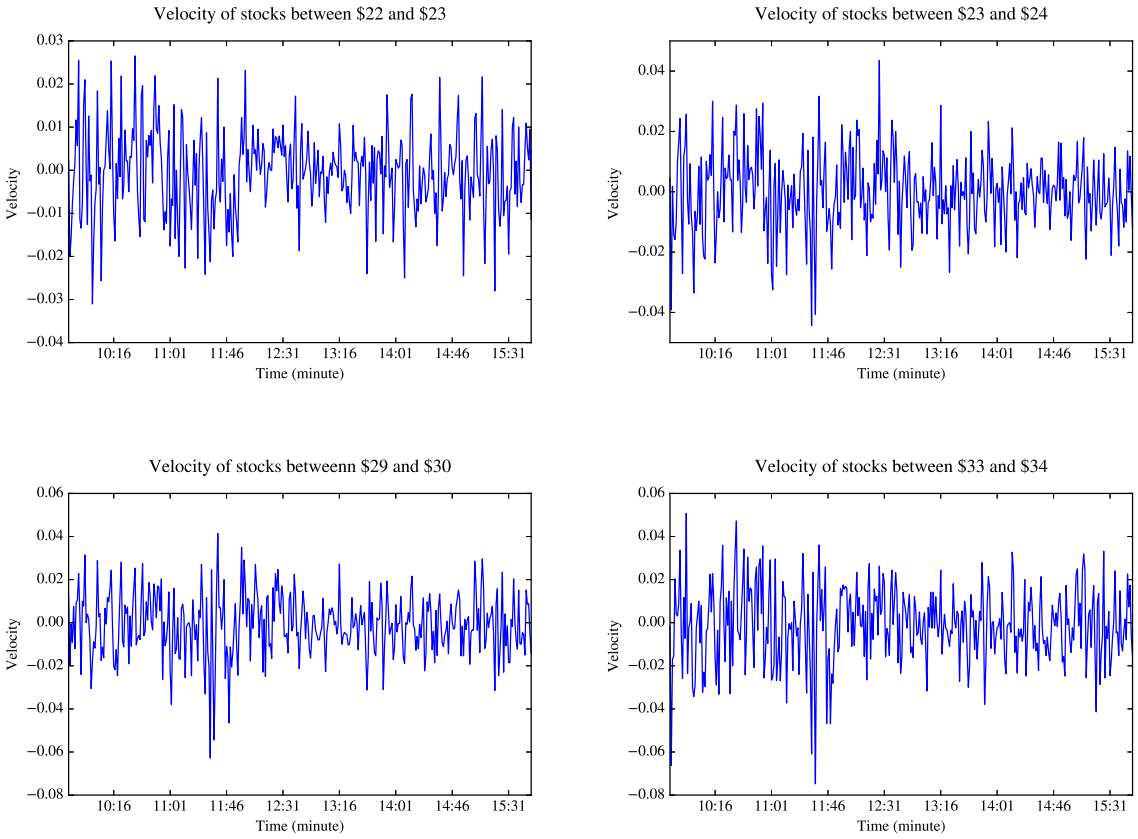


Figure 3.10: Velocity Plots

Three of the plots show that there was an unusual dip (or a crash) between 11:00 a.m. and 11:45 a.m. on May 18, 2018, which was the reported time for Facebook public offering. Though the instability was caused by one stock, the glitches caused by that massive offering on the NASDAQ system created chaos across the market. We can see a dip around the crash time in different price ranges. This means that the crash propagated from one price range to another, supporting the wave-like phenomenon of equity markets.

The prediction results for the flux and density are shown in the tables and figures below.

Figure 3.11-3.12 show one-ahead predictions of the density and flux for four price ranges. The figures show that the model is able to capture the fluctuations during normal and high volatility periods (i.e., around 11:30 am). We also see that the observations are within the predicted confidence bound.

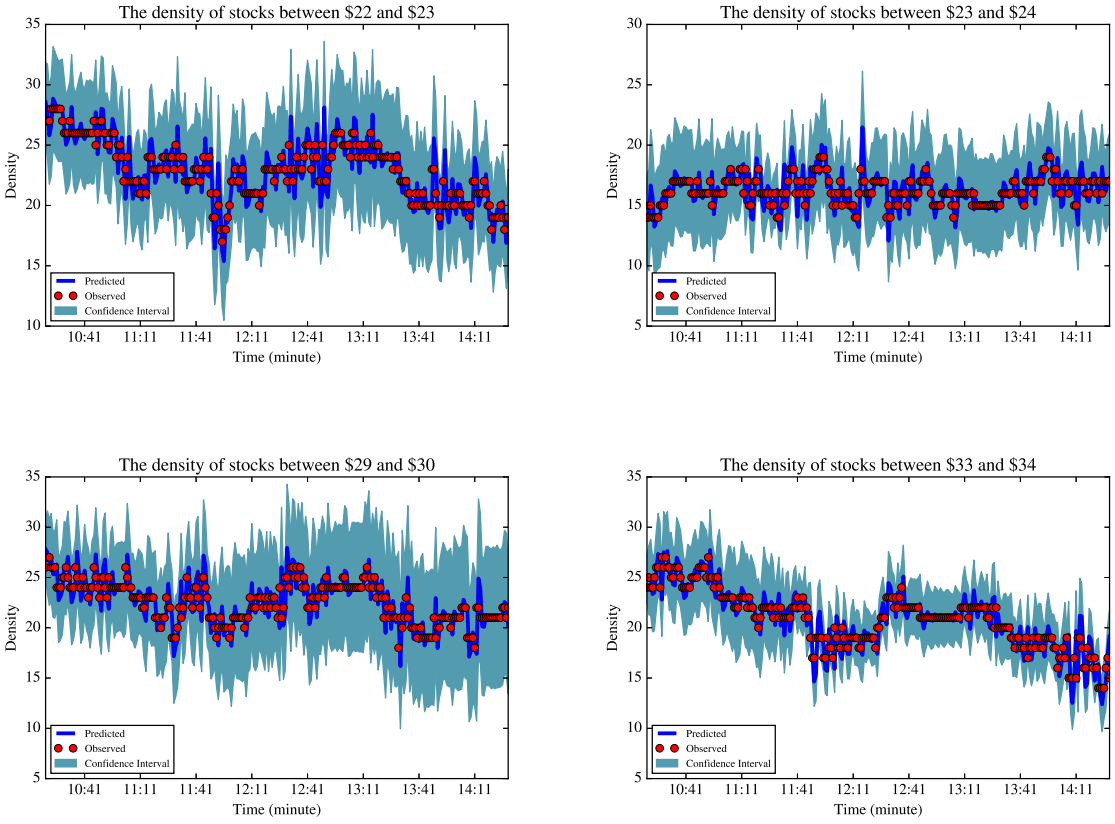


Figure 3.11: One-Step Ahead Prediction of Density

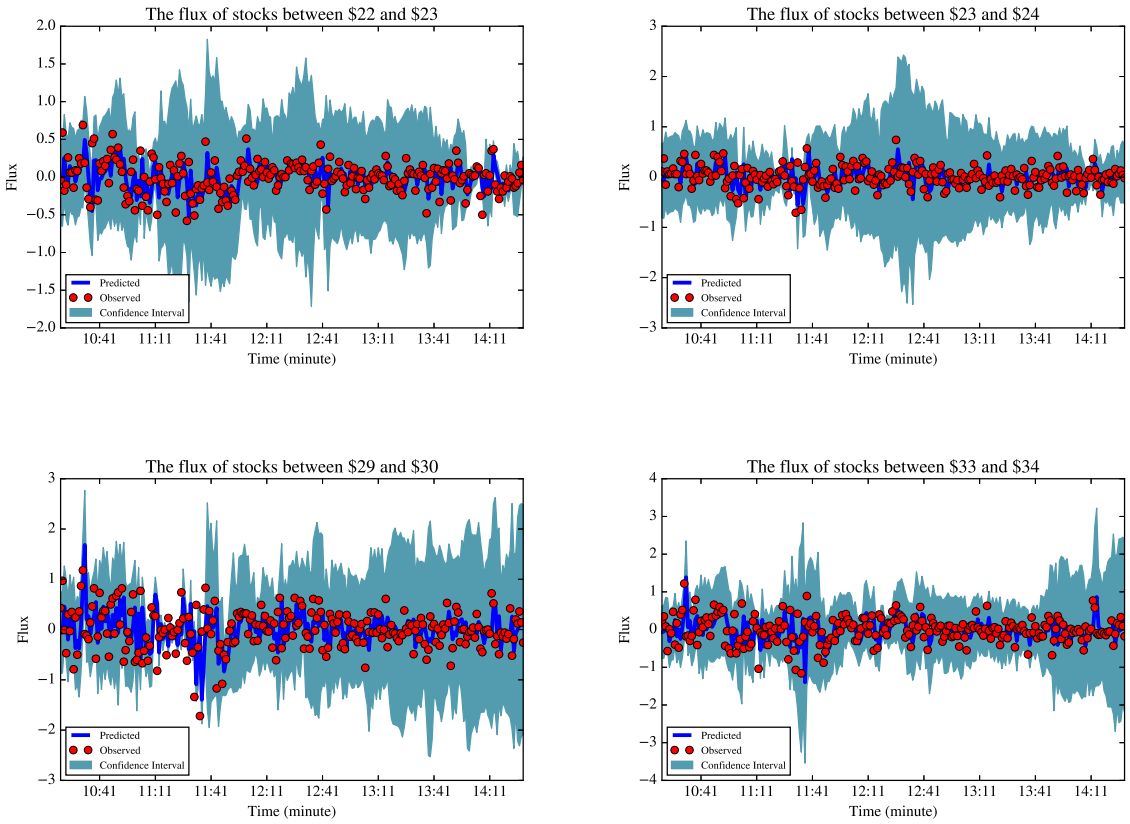


Figure 3.12: One-Step Ahead Prediction of Flux

The tables below show that the implantation of the combination of Lax-Wendroff and UKL methods results in the minimum RMSE. Note that the errors from using other methods are comparable.

Model		RMSE	MSE
PDEs Solver	Stochastic Filtering Technique		
HLL	Dual UKL	1.24	0.91
HLL	Dual EnKL	2.68	1.98
Lax-Wendroff	Dual UKL	1.22	0.91
Lax-Wendroff	Dual EnKL	2.94	2.17

Table 3.3: Average Performance of the Proposed Algorithms in Predicting the Density of Stocks

Model		RMSE	MSE
PDEs Solver	Stochastic Filtering Technique		
HLL	Dual UKL	0.29	0.22
HLL	Dual EnKL	0.29	0.22
Lax-Wendroff	Dual UKL	0.29	0.22
Lax-Wendroff	Dual EnKL	0.29	0.22

Table 3.4: Average Performance of the Proposed Algorithms in Predicting the Flux of Stocks

Figure 3.13 illustrates the predicted pressure. It is evident that the pressure is more informative about unusual changes in the market, as seen in the plots around 11:30 a.m. The four figures have a common peak around that time.

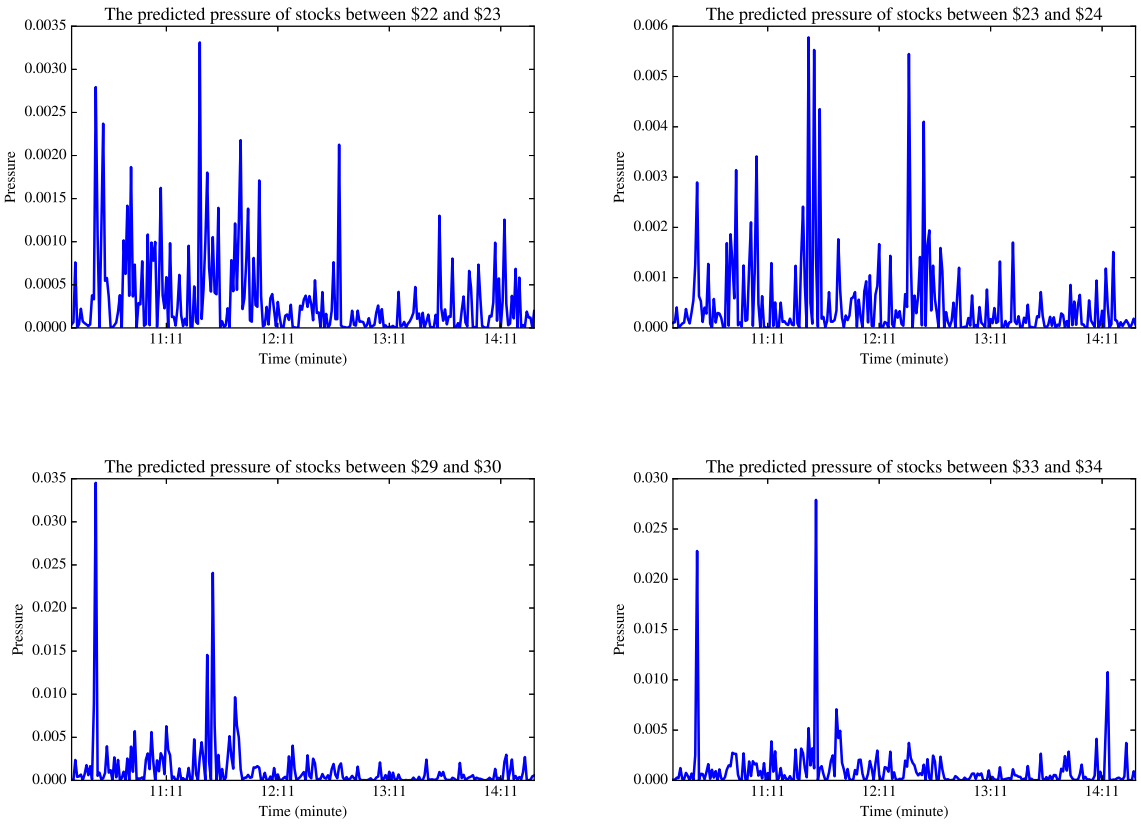


Figure 3.13: One-Step Ahead Prediction of Pressure

3.5.5 Sensitivity Analysis on α

In this section, we analyze the robustness of the MEM under different values of α. In the theoretical and implementation sections, the value of α was assumed to be a constant parameter and between 0 and 1. More specifically, we picked $\alpha = 0.3$. To compare the performance of the model under different values of α, we compute the difference between the prediction results when $\alpha = 0.3$ and other values between 0 and 1. Tables 3.5 and 3.6 show that the results are almost identical for several values of $\alpha = 0.3$, indicating that the MEM is robust to changes in α in the range between 0 and 1.

Prediction Average Difference (base case is α=0.3)		
α	Flash Crash Day	Facebook's IPO
0.1	4.97 $\times 10^{-6}$	4.7 $\times 10^{-6}$
0.6	-7.42 $\times 10^{-6}$	-5.5 $\times 10^{-6}$
0.8	-1.24 $\times 10^{-5}$	-8.17 $\times 10^{-6}$

Table 3.5: The Difference in the Prediction Results (RMSE) of Density between $\alpha = 0.3$ and other Values of α

Prediction Average Difference (base case is $\alpha = 0.3$)		
α	Flash Crash Day	Facebook's IPO
0.1	2.21 $\times 10^{-6}$	1.42 $\times 10^{-5}$
0.6	-3.29 $\times 10^{-6}$	-2.26 $\times 10^{-5}$
0.8	-5.49 $\times 10^{-6}$	-3.84 $\times 10^{-5}$

Table 3.6: The Difference in the Prediction Results (RMSE) of Velocity between $\alpha = 0.3$ and other Values of α

3.6 Conclusion

This chapter presents a new narrative to flash crashes. During normal periods, the markets are undergoing a lot of weak shocks that we believe are absorbed by market participants. During flash crashes (or macroscopic crashes), at early stages, a strong shock builds up and hits the market aggressively. Then, said shock, propagates to the whole market and eventually is dumped down. Our study of shock speed and strength establishes strong evidence connecting shocks to crashes. The analysis is established through the access of high frequency data. In this chapter, the structure of the MEM model was investigated for the first time. The conditions in which shock or rarefaction waves occur were highlighted. In the equity markets, shocks are formed when stock prices, captured by the velocity of stocks, experience a sharp decline. A rarefaction wave is a solution in which stocks velocity transition smoothly (to some extent) from one state to another. We proposed a solution methodology of the MEM model. This approach is novel, combining a PDEs solver and stochastic filtering techniques. The solution method is adaptive and capable of updating the estimates to reflect any changes in market conditions. The results of implementing the proposed approach on two datasets illustrated that pressure is more sensitive to crashes. Thus, it can be used as a predictive tool for crashes and to provide early alert for abnormalities in the market. The prediction results of density and flux for several price ranges were within the predicted confidence bound.

3.7 Technical Proofs

3.7.1 Derivation of 3.20

$$Q_m = -\frac{Q_*\rho_m}{(\alpha\rho_m - (1+\alpha)\rho_*)} + \sqrt{Q_*^2\Big(\frac{\rho_m\rho_* + (1+\alpha)\rho_m(\rho_m-\rho_*)}{\rho_*(\alpha\rho_m-(1+\alpha)\rho_*)} + \frac{\rho_m^2}{(\alpha\rho_m-(1+\alpha)\rho_*)^2}\Big)} \tag{3.86}$$

$$= -\frac{Q_*\rho_m}{(\alpha\rho_m - (1+\alpha)\rho_*)} \tag{3.87}$$

$$+ \sqrt{Q_*^2\Big(\frac{\Big(\rho_m\rho_* + (1+\alpha)\rho_m(\rho_m-\rho_*)\Big)(\alpha\rho_m-(1+\alpha)\rho_*)}{\rho_*(\alpha\rho_m-(1+\alpha)\rho_*)^2} + \frac{\rho_m^2}{(\alpha\rho_m-(1+\alpha)\rho_*)^2}\Big)}, \tag{3.88}$$

$$= -\frac{Q_*\rho_m}{(\alpha\rho_m - (1+\alpha)\rho_*)} \tag{3.89}$$

$$\pm \frac{Q_*}{(\alpha\rho_m-(1+\alpha)\rho_*)}\sqrt{\frac{\Big(\rho_m\rho_* + (1+\alpha)\rho_m(\rho_m-\rho_*)\Big)(\alpha\rho_m-(1+\alpha)\rho_*)}{\rho_*} + \rho_m^2}, \tag{3.90}$$

$$= -\frac{Q_*\rho_m}{(\alpha\rho_m - (1+\alpha)\rho_*)} \tag{3.91}$$

$$\pm \frac{Q_*}{(\alpha\rho_m-(1+\alpha)\rho_*)}\sqrt{\frac{\rho_m^3(\alpha+\alpha^2) - 2\rho_m^2\rho_* + \rho_*^2\rho_m(\alpha+\alpha^2) - \rho_m^2\rho_*}{\rho_*} + \rho_m^2}, \tag{3.92}$$

$$= -\frac{Q_*\rho_m}{(\alpha\rho_m - (1+\alpha)\rho_*)} \pm \frac{Q_*}{(\alpha\rho_m-(1+\alpha)\rho_*)}\sqrt{\frac{\rho_m^4(\alpha+\alpha^2) - 2\rho_m^3\rho_* + \rho_*^2\rho_m^2(\alpha+\alpha^2)}{\rho_m\rho_*}}, \tag{3.93}$$

$$= -\frac{Q_*\rho_m}{(\alpha\rho_m - (1+\alpha)\rho_*)} \pm \frac{Q_*}{(\alpha\rho_m-(1+\alpha)\rho_*)}\sqrt{(\rho_m^2-\rho_m\rho_*)^2\frac{(\alpha+\alpha^2)}{\rho_m\rho_*}}, \tag{3.94}$$

$$= Q_*\frac{1 \pm (\rho_m-\rho_*)\sqrt{\frac{\alpha^2+\alpha}{\rho_m\rho_*}}}{1-\frac{(\rho_m-\rho_*)}{\rho_m}(1+\alpha)}. \tag{3.95}$$

3.7.2 Proof of Theorem 3.10

Proof. It follows from the entropy condition that

$$(1+\alpha-\sqrt{\alpha^2+\alpha})\,\frac{Q_R}{\rho_R} < \frac{Q_R-Q_L}{\rho_R-\rho_L} < (1+\alpha-\sqrt{\alpha^2+\alpha})\,\frac{Q_L}{\rho_L}, \tag{3.96}$$

$$(1+\alpha-\sqrt{\alpha^2+\alpha})\,\frac{Q_R}{\rho_R} - \frac{Q_R-Q_L}{\rho_R-\rho_L} < 0 < (1+\alpha-\sqrt{\alpha^2+\alpha})\,\frac{Q_L}{\rho_L} - \frac{Q_R-Q_L}{\rho_R-\rho_L}, \tag{3.97}$$

after a few steps, we can see that

$$\frac{Q_L\rho_R - Q_R\rho_R}{\rho_R-\rho_L} > -(\alpha-\sqrt{\alpha^2+\alpha})Q_L, \tag{3.98}$$

$$\frac{Q_L\rho_R - Q_R\rho_R}{\rho_R-\rho_L} < -(\alpha-\sqrt{\alpha^2+\alpha})Q_R, \tag{3.99}$$

since by definition $(\alpha-\sqrt{\alpha^2+\alpha}) < 0$, we obtain that for 1-shock $Q_R > Q_L$ and take the minus sign in (3.20). The first family of shock curve $S1$ follows easily by replacing Q_m and Q_* in (3.20) by Q_R and Q_L, respectively, and taking the minus sign. The shock speed s_1 is obtained by substituting Q_R in (3.17) by (3.10).

To show that $\rho_R > \rho_L$ when $Q_L > 0$, we discuss two cases

1. If the denominator in (3.10) is negative: $1-\frac{(\rho_R-\rho_L)}{\rho_R}(1+\alpha) < 0$, hence $Q_R > Q_L$ if $0 > 1-\frac{(\rho_R-\rho_L)}{\rho_R}(1+\alpha) > 1-(\rho_R-\rho_L)\sqrt{\frac{\alpha^2+\alpha}{\rho_R\rho_L}}$ or $\rho_R > \rho_L$. Otherwise, we get a contradiction.

2. If the denominator in (3.10) is positive: $1-\frac{(\rho_R-\rho_L)}{\rho_R}(1+\alpha) > 0$, hence $Q_R > Q_L$ if $1-(\rho_R-\rho_L)\sqrt{\frac{\alpha^2+\alpha}{\rho_R\rho_L}} > 1-\frac{(\rho_R-\rho_L)}{\rho_R}(1+\alpha) > 0$ or $\rho_R > \rho_L$. Otherwise, we get a contradiction.

To show that $\rho_R < \rho_L$ when $Q_L < 0$, we discuss two cases

1. If the denominator in (3.10) is negative: $1-\frac{(\rho_R-\rho_L)}{\rho_R}(1+\alpha) < 0$, hence $Q_R > Q_L$ if $0 > 1-(\rho_R-\rho_L)\sqrt{\frac{\alpha^2+\alpha}{\rho_R\rho_L}} > 1-\frac{(\rho_R-\rho_L)}{\rho_R}(1+\alpha)$ or $\rho_R < \rho_L$. Otherwise, we get a contradiction.

2. If the denominator in (3.10) is positive: $1 - \frac{(\rho_R - \rho_L)}{\rho_R}(1+\alpha) > 0$, hence $Q_R > Q_L$ if $0 < 1 - (\rho_R - \rho_L)\sqrt{\frac{\alpha^2+\alpha}{\rho_R \rho_L}} < 1 - \frac{(\rho_R - \rho_L)}{\rho_R}(1+\alpha)$ or $\rho_R < \rho_L$. Otherwise, we get a contradiction.

□

3.7.3 Proof of Theorem 3.11

Proof. This proof follows easily from the previous proof, but instead we use the fact that $(\alpha + \sqrt{\alpha^2+\alpha}) > 0$. □

3.7.4 Proof of Proposition 3.12

Proof. We only show the proof the first case as the other cases follow similarly. It follows from the proof of Theorem 3.10 that if $1 - \frac{(\rho_R - \rho_L)}{\rho_R}(1+\alpha) < 0$,which is equivalent to say that $\frac{\rho_L}{\rho_R} < \frac{\alpha}{1+\alpha}$, and $Q_L > 0$,then $\rho_R > \rho_L$ and $\frac{(\rho_R - \rho_L)}{\rho_R}(1+\alpha) < (\rho_R - \rho_L)\sqrt{\frac{\alpha^2+\alpha}{\rho_R \rho_L}}$, which also means that $\frac{1+\alpha}{\rho_R} < \sqrt{\frac{\alpha^2+\alpha}{\rho_R \rho_L}}$ since $\rho_R > \rho_L$. Thus, we obtain from (3.25) and (3.28) that $s_1 > 0$ and $s_2 < 0$. □

3.7.5 Proof of Theorem 3.13

Proof. We determine explicitly the function $\tilde{q}(x/t)$ by first determining $\mu(\zeta)$ in (3.33). We differentiate (3.35) with respect to ζ to obtain

$$1 = \nabla\lambda_i(\tilde{q}(\zeta)) \cdot \tilde{q}'(\zeta) \tag{3.100}$$

$$= \mu(\zeta)\nabla\lambda_i(\tilde{q}(\zeta)) \cdot r_i(\tilde{q}) \tag{3.101}$$

where the second equality follows from using (3.33), and thus

$$\mu(\zeta) = \frac{1}{\nabla\lambda_i(\tilde{q}(\zeta)) \cdot r_i(\tilde{q})} \tag{3.102}$$

Note that the denominator of (3.102) is not zero due to the characteristic fields being genuinely nonlinear (Theorem 3.2).

Using (3.102) in (3.33), and hence

$$\tilde{q}'(\zeta) = \frac{r_i(\tilde{q})}{\nabla\lambda_i(\tilde{q}(\zeta)) \cdot r_i(\tilde{q})} \tag{3.103}$$

which is a system of ordinary differential equations for $\tilde{q}(\zeta)$ when $\zeta_1 \leq \zeta \leq \zeta_2$ with a initial condition $\tilde{q}(\zeta_1) = q_L$. For 1-rarefaction of the MEM model , the system of ODEs is given by

$$\rho'(\zeta) = \frac{\rho^2(\zeta)}{Q(\zeta)(1+\alpha-\sqrt{\alpha^2+\alpha})\Big((1+\alpha-\sqrt{\alpha^2+\alpha})-1\Big)}, \;\; \rho(\zeta_1) = \rho_L \tag{3.104}$$

$$Q'(\zeta) = \frac{\rho(\zeta)}{\Big((1+\alpha-\sqrt{\alpha^2+\alpha})-1\Big)}, \;\; Q(\zeta_1) = Q_L \tag{3.105}$$

Divide the previous two equation to obtain

$$\frac{dQ(\zeta)}{d\rho(\zeta)} = (1+\alpha-\sqrt{\alpha^2+\alpha})\frac{Q(\zeta)}{\rho(\zeta)} \tag{3.106}$$

$$\tag{3.107}$$

rearranging the terms, which yields

$$\frac{dQ(\zeta)}{Q(\zeta)} = (1+\alpha-\sqrt{\alpha^2+\alpha})\frac{d\rho(\zeta)}{\rho(\zeta)} \tag{3.108}$$

hence, we eliminate ζ by solving for Q as a function of ρ and obtain

$$\ln\left(\frac{Q(\rho)}{Q_L}\right) = (1+\alpha-\sqrt{\alpha^2+\alpha})\ \ln\left(\frac{\rho}{\rho_L}\right), \tag{3.109}$$

thus, for a 1-rarefaction wave connecting q_L to q_R, the first part of the theorem holds

$$Q_R = Q_L\Big(\frac{\rho_R}{\rho_L}\Big)^{1+\alpha-\sqrt{\alpha^2+\alpha}}. \tag{3.110}$$

Since $(1+\alpha-\sqrt{\alpha^2+\alpha})$ is between zero and one (by definition), and if Q_L ¿0,

$$Q_R = Q_L\Big(\frac{\rho_R}{\rho_L}\Big)^{1+\alpha-\sqrt{\alpha^2+\alpha}} > Q_L\frac{\rho_R}{\rho_L}. \tag{3.111}$$

The previous inequality follows from the existence condition of a rarefaction wave, which is given by

$$\lambda_1(q_L) = (1+\alpha-\sqrt{\alpha^2+\alpha})\cdot\frac{Q_L}{\rho_L} < \lambda_1(q_R) = (1+\alpha-\sqrt{\alpha^2+\alpha})\cdot\frac{Q_R}{\rho_R} \tag{3.112}$$

Hence,

$$\rho_R^{\alpha-\sqrt{\alpha^2+\alpha}} > \rho_L^{\alpha-\sqrt{\alpha^2+\alpha}} \tag{3.113}$$

which means that $\rho_R < \rho_L$. Also, we can easily see that $\rho_R > \rho_L$ when $Q_L < 0$. This completes the proof. □

3.7.6 Proof of Theorem 3.14

Proof. This proof follows easily from the previous proof, but instead we use the fact that $(\alpha+\sqrt{\alpha^2+\alpha}) > 0$. □

Chapter 4

Abnormality Detection in Equity Markets: A Macroscopic Hybrid Reservoir Computing Approach

4.1 Introduction

In recent years, algorithmic trading in the financial markets has significantly increased and replaced open outcry mechanisms [61]. Algorithm traders use computers to follow specific instructions to exploit profitable opportunities as they arise. One type of algorithmic trading is High Frequency, which uses powerful and fast computers, and complex algorithms to yield high returns. With the spread of these trading behaviors, we have also seen the emergence of new market anomalies such as flash crashes. Indeed, empirical studies have confirmed the impact of high frequency trading on stock prices, creating market imbalance and instabilities [84]. Previous research on market anomalies has, primarily, focused only on those that violated the market efficiency hypothesis [106]. Yet, in this era of fast computing, anomalies can also occur in response to complex transactions and glitches. Even though technological development in trading practices is important for the market to advance, there remains a need for a monitoring tool that can detect abnormal activities.

The existing research into predicting financial anomalies [1] has mainly focused on major crisis forecasting. These forecasting methods include classical models and machine learning techniques. One of the classical approaches is to use macroeconomics indicators to predict financial stress. In [118], a Bayesian Model Averaging (BMA) is used to identify variables with high predictive power for financial stress in 25 of the Organization for Economic Cooperation and Development (OECD) countries. Despite the success of BMA in accounting for model uncertainty, the authors of [118] reported that the BMA was not able to provide good predictions for financial stress. Other studies have used traditional econometric models to forecast economic crises. In [43], for instance, early warning indicators were identified using a linear model known as Auto Regressing Moving Average (ARMA). A different line of research in forecasting financial crises is the implementation of machine learning techniques. In this line of research, early detection systems were developed using state of the art machine learning techniques such as Artificial Neural Network (ANN), decision trees, and logistic regression. Of these three methods, ANN has yielded the best performance. Tree-based machine learning techniques have also been investigated in the literature. Boosted Regression Trees (BRT) were used in [33] to select indicators for predicting recessions, showing that the short-term interest rate and the term spread were the leading indicators. The previous research in forecasting tools have focused on high intensity crashes. However, little research has investigated the *medium intensity crashes* that can occur during a given trading day. We consider these crashes as anomalies that are important to identify.

Decision makers need a powerful tool to detect and discover early market anomalies in order to reduce their effect. At present, abnormality detection methods using machine learning techniques in financial applications fall into six main categories: regression, clustering, prediction, outlier detection, visualization, and classification methods [93]. Clustering methods don't require the data to be labeled. By employing a similarity measure, clustering aims to group similar data together. Classification methods are best suited for labeled data in which

[1] The terms "abnormalities" and "anomalies" are used interchangeably throughout this chapter.

a classification model is trained to identify class boundaries. The trained model is then used to classify future instances. In this research, we are particularly focused on the application of classification algorithms to detect market anomalies, specifically medium intensity crashes.

Medium intensity crashes are relatively scarce compared to regular market activity, making them a underrepresented minority. This class imbalance is a common problem in real-world datasets and many classification techniques. The imbalance hinders the ability of classification algorithms to learn the decision boundaries correctly [130]. Previous research on class imbalance problems has typically focused on approaches at the data and algorithmic levels. Approaches at the data level include re-sampling methods that adjust the data class distribution [113]. Examples of sampling methods encompass: oversampling, undersampling, and threshold moving methods [23]. At the algorithmic level, existing approaches involve cost-sensitive methods [113]. These methods do not attempt to balance the distribution of classes. Instead, cost metrics are introduced in the learning procedure to minimize misclassifying samples. The class imbalance problem using cost-sensitivity has been addressed extensively in the literature [71]. Also, a plethora of literature has shown, empirically, that cost-sensitive approaches yield satisfactory results. So far, however, there has been little research on the class imbalance learning of Reservoir Computing (RC), which is a relatively new classification method that processes time-dependent inputs efficiently. Furthermore, to the best of our knowledge, this research is the first to use RC in predicting market anomalies.

This study aims to develop a method for identifying abnormalities (specifically medium intensity crashes) in equity markets using RC. RC is a recurrent neural network in which most of the network parameters are generated randomly, with others being optimized or trained [87]. We propose a hybrid cost-sensitive RC model to classify medium frequency trading into normal and abnormal activity. The proposed model utilizes a hybrid set of inputs composed of multivariate time series of macroscopic variables, and features obtained from the Macroscopic Equity Markets Model [2]. We find that the proposed hybrid model is able to detect abnormalities with high accuracy, and low false positive and false negative

rates.

This chapter is organized as follows. Section 2 reviews recent literature on classification methods. The problem description is presented in Section 3. The model is proposed in section 4. The results of implementing the proposed model are discussed in section 5. Section 6 concludes the chapter.

4.2 Related Works

A series of local crashes can give rise to a global financial crisis due to the connectivity of financial markets. A crash originating in one country might propagate to other countries. For instance, the sub-prime crises in the United States propagated to a number of countries in Europe, creating a sovereign debt crisis. This event is an example of the turbulence effect of financial crises. The turbulence effect has been documented in the literature, warning the need for an early crash detection method [7, 44].

The financial stress (or crash) that we study in this research is unique in that it affects most of equity markets on a specific day, then vanishes. We tackle the problem of identifying medium intensity crashes as a classification problem of multivariate time series and only review the related works that are centered around this approach. The classification problem of multivariate time series has been investigated in several domains, including: health application [69, 19], action recognition [59], and civil engineering [20]. This problem consists of classifying each time series to a specific class or label. The classification problem has been approached from different directions, ranging from methods based on similarity measures (e.g. Weighted Dynamic Time Wrapping) to ensemble classifiers [5]. In our research, we focus on ANNs, in particular Recurrent Neural Networks (RNNs), to classify multivariate time series, which are computational methods to process time series data that exhibit time dependency [54].

RNNs are computational models with a special structure. Similar to the Feed-Forward Network (FFN), an RNN has inputs, links to hidden layers, and an output. What makes an

RNN different from FFN is that it includes a feedback link from the output to the network itself. An RNN is considered a dynamical system, while FFN is a functional approximation model. RNNs have shown highly promising results, which are attributed to being universal approximators of dynamical systems [48]. A further advantage of RNN is that it closely resembles the brain structure by having feedback connections. This feature, in particular, makes RNNs more successful in solving many real-life applications when feedback information is important.

Despite the advantages of RNNs, they suffer from major drawbacks. RNNs are challenging to train (training is usually done by gradient-decent type methods), and the convergence of the gradient information is not guaranteed [34]. The second shortcoming is that RNN is computationally very expensive. For instance, the training time for a single parameter update is long. The third shortcoming is that the network usually fails to learn long-range memory which is caused by the exploding of the gradient over time [9]. This problem has been addressed in the literature, resulting in the introduction of the Long-Short Term memory networks [50].

Reservoir Computing (RC) is a class of RNNs whose internal connections are randomly generated [60, 102], which offers a faster training speed. Though the architecture of RC is simple, it provides a rich representation of the input through what is called the reservoir. This rich representation is very helpful in solving real-life applications. The only aspect of RC that requires training are the connections between the reservoir and the final output (called readout weights) [87, 104]. Given its advantages, RC has been applied successfully to many problems, spanning from time series forecasting [108, 32, 10], to speech analysis [117]. Also, a variety of research has shown that RC is an efficient classifier for multivariate time series [114, 89].

To improve the performance of RC, the model space classification was introduced in [24]. This method aims to use the readout weights of RC to represent the multivariate time series. To model space classification, the authors of [3] have found that the classification

of time series data, using only the last time step to represent the series, has yielded poor results. Furthermore, they have highlighted that the time step wise classification (called the reservoir space) and the classification using the trained connections (called the model space) have resulted in good accuracy. Other efforts to improve the performance of RC resulted in a method that ignores the training of the readout weights and applies principle components analysis to the reservoir space [98].

The drawback of the previous methods is that the reservoir accounts only for the information obtained by nonlinear transformation of the input. In doing so, the model based information coming from a physics-based model is ignored. To overcome this shortcoming, we extend RC to process information from a specific physics-based model called the Macroscopic Equity Markets (MEM) model [2] and allow all the information to be processed in the reservoir and the output.

Our work contributes to the literature by proposing a cost-sensitive RC for the imbalance classification problem. Our work offers a novel perspective on forecasting crashes. This is a new solution to the medium crashes prediction method that has not been reported in the literature. Another key novelty of our approach consists in the construction of a macroscopic hybrid classification method based on RC. To this end, the contribution of our new approach lies in the following areas. First, we use multivariate time series composed of macroscopic variables to label instances (as normal and abnormal) during a trading day. The abnormality instances are linked to high volatility levels. Second, a cost-sensitive RC is proposed to tackle the imbalance classification problem. Third, the proposed RC combines macroscopic financial variables obtained from our previous research [2] and features obtained from MEM model. Finally, we focus on unique datasets that covers more than fifteen days from the last ten years in which medium crashes have been reported.

Notations: Throughout the chapter, we use bold and small letters to represent vectors, and bold and capital letters to represent matrices. For example: $\mathbf{x}$ is a vector while $\mathbf{X}$ is a matrix. For a vector $\mathbf{x} \in \mathbb{R}^n$, the L_2 norm is given by $\|\mathbf{x}\|_2 := \sqrt{x_1^2 + \cdots + x_n^2}$. The

vector e_j denotes an all zero vector with 1 in the jth element. The symbol $\oplus$ denotes the concatenation of two vectors.

4.3 Problem Description

Suppose that we have d trading days with the empirical data $\{\mathbf{x}(t), y_{target}(t+1)\}^d$, where $\mathbf{x}(t) \in \mathbb{R}^M$ is the given raw data input at time t or feature and $y_{target}(t) = \{0, 1\}$ is the true class or output at time t, where 0 corresponds to a normal day and 1 to an abnormal trading day. The empirical data is available for $t = 0, \cdots, T-1$, indicating that the input and output are collected in discrete time with a time step $\Delta t = 1$. The classification task is divided into two stages: training and testing stages. In the training stage, the classification task aims to learn (or fit) a classifier $y(t+1) = f(\mathbf{x}(t), \mathbf{x}(t-1), \cdots, \mathbf{x}(0); \mathbf{w})$, which is parameterized by $\mathbf{w}$, using the empirical data during $0 \leq t \leq T-1$, where the classifier $f : \mathbb{R}^N \to \{0, 1\}$, such that a loss function $l(y, y_{target})$ is minimized. In the testing stage, the fitted classifier is used to make predication about a newly encountered trading day with a new $\mathbf{x}(t)$.

Often, feature extraction is implemented to extract more properties of $\mathbf{x}(t)$ which might be more relevant in the classification task. For an input $\mathbf{x}(t)$, the feature vector is represented as $\mathbf{u}(t) = \phi(\mathbf{x}(t))$, where ϕ is a feature extraction function . The feature vector allows the rich representation about the input to be captured; thus improving the learning process. The classifier can be rewritten as $f\Big(\mathbf{u}(t-1), \mathbf{u}(t-2), \cdots; \mathbf{w}\Big)$. With the previous representation of the classifier, the classification problem centers on learning the classifier f so that the minimum loss is achieved.

4.4 Methods

4.4.1 Preliminaries

In order to understand the structure of RC, it is essential to cover a few concepts in Artificial Neural Networks (ANNs) (for a comprehensive review, see [53]). Generally speaking, an ANN is a computational model that is biologically inspired by the human brain. An ANN consists of three main layers: input, hidden and output layers. The network receives the data through the input layer. In the hidden layer, computational unites (called neurons) apply a transformation to the input, which then is sent to the output layer. Different architectures for neural networks have been proposed in the literature. We focus here on two main types: Feed Forward Neural Network (FFN) and Recurrent Neural Network (RNN).

4.4.1.1 Feed Forward Neural Netowrk (FFN)

A FFN approximates a function $y_{target} = f_{target}$. The goal of the network is to define a map $y = f(\mathbf{x}; \mathbf{W})$ and learn the value of the weight parameters $\mathbf{W}$ such that the learned function is the best approximation applicable. In feed forward networks, the information flows from the input being $\mathbf{x}$, then through computational nodes that are used to define the function f, and lastly to the output y. In these types of networks, there are no feedback connections from the output to the network. Networks with feedbacks are referred to as RNN.

A feed forward network is composed of different functions. For instance, a function $f(\mathbf{x})$ can be composed of other functions that are connected in a chain, $f(\mathbf{x}) = f_3(f_2(f_1(\mathbf{x})))$. In that functional structure of neural networks, f_1 is called the first layer and, so on. The number of layers defines the depth of the model. The middle layers are called hidden layers. The last layer is the output layer. Each layer consist of many units called neurons. Each neuron is a function that transforms a vector to a scalar. We illustrate a feed forward network through a simple network, as shown in Figure 4.1.

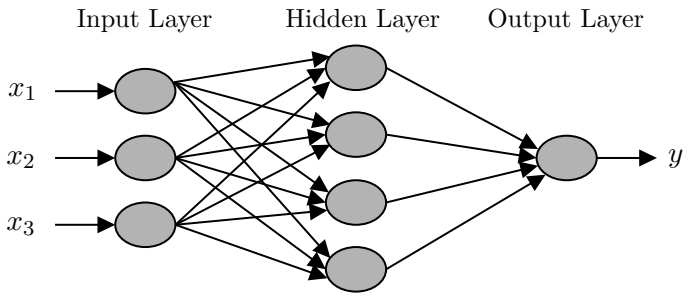


Figure 4.1: Feed Forward Neural Network

In Figure 4.1, the first layers represent a vector of hidden units $\mathbf{h}$. This layer is used to compute a vector $f_1(\mathbf{x}; \mathbf{W}_1, \mathbf{b}_1)$, where $\mathbf{W}_1$ and $\mathbf{b}_1$ are the weights matrix (i.e., connections from units in the input layer to unites in the first layer) and bias vector for the first layer. In the hidden layer, the function f_1 is called an activation function and is applied to each unit in the layer. A hidden unit can be represented as $h_i = g(\mathbf{W}_{:,i}\mathbf{x} + \mathbf{b}_{1i})$, where h_i is the hidden unit i, $\mathbf{W}_{:,i}$ is the weights vector of the connections from the input layer to unit i, and $\mathbf{b}_{1i}$ is the bias of unit i in the first layer. There are many choices for the function g that have been used in the literature, most notably: Rectified linear unit and tanh functions [53]. The value of the function f_1 is used as the input for the second layer (i.e., the output layer). Thus, the network's final function is the chain of two functions: the hidden layer function $\mathbf{h} = f_1$ and the output layer function $y = f_2(\mathbf{h}, \mathbf{w}_2, b_2)$, where $\mathbf{w}_2$ is the weight vector (i.e., connections from the first layer to the output) and b_2 is the bias. The final form of the model is $y = f_2(f_1(\mathbf{x}))$. The function f_2 depends on the task, which can be regression or classification.

The training of neural networks aim to sequentially update the network weights and biases to minimize a loss function $l(y, y_{target})$. The loss function l penalizes the deviation of y_{target} from y. The training process is usually performed by the back-propagation algorithm [53]. This algorithm calculates the derivative of the loss function l with respect to the parameters in the network. The parameters are updated by the gradient descent algorithm.

4.4.1.2 Recurrent Neural Networks (RNNs)

Recurrent Neural Networks (RNNs) are a particular category of ANNs for modeling sequential and time series data. A RNN is a feed forward neural network characterized by including information from adjacent time steps, featuring the notion of time to the network. This feature allows the network to learn the sequential/temporal dependencies in the data. A basic structure of RNN is illustrated in Figure 4.2

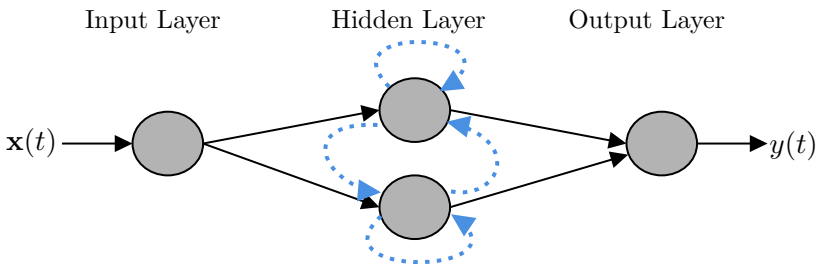


Figure 4.2: Recurrent Neural Network: Blue dashed lines indicate hidden-to-hidden recurrent connections from time $t-1$ to time t

Let the input of the RNN be a vector $\mathbf{x}(t)$ indexed by time t, starting from 0 to T. At time t, nodes in the hidden layer receive connections from the current input vector $\mathbf{u}(t)$ and from hidden nodes from $t-1$. The hidden layer at time t can be represented by a vector $\mathbf{h}(t) = g(\mathbf{W}_{uh}\mathbf{x}(t) + \mathbf{W}_{hh}\mathbf{h}(t-1) + \mathbf{b}_h)$, where b_h is the bias vector, $\mathbf{W}_{uh}$ is the weight matrix for the input-to-hidden connections, and $\mathbf{W}_{hh}$ is the weight matrix for the hidden-to-hidden connections. The output $y(t) = f_2(\mathbf{h}(t), \mathbf{w}_2, b_2)$, is a function of the hidden nodes $\mathbf{h}(t)$ at time t, making the input $\mathbf{u}(t)$ at time $t-1$ influences the output $y(t)$ at time t by the recurrent connections, and so on. Figure 4.2 shows a simple structure of RNNs. This network can be extended to include many layers (deep network). RNNs are commonly trained by BackPropagation Through Time (BPTT). BPTT is an extension to the gradient-based algorithm that is used to train FNN.

Though RNNs are a promising approach for modeling complex time series data, training of RNNs by gradient-decent based approaches is difficult [87]. In the face of RNNs challenges,

two new approaches, namely Liquid State Machines (LSMs) [90] and Echo State Networks (ESNs) [60], were proposed independently to overcome these challenges. Both approaches are referred to as RC. In this chapter, we study ESN, which is a large network with many randomly generated interconnections between nodes in the hidden layers.

4.4.2 Classification based on Reservoir Computing

In this section, we propose an RC architecture to classify high frequency financial data. In Figure 4.3, we show a basic structure of RC.

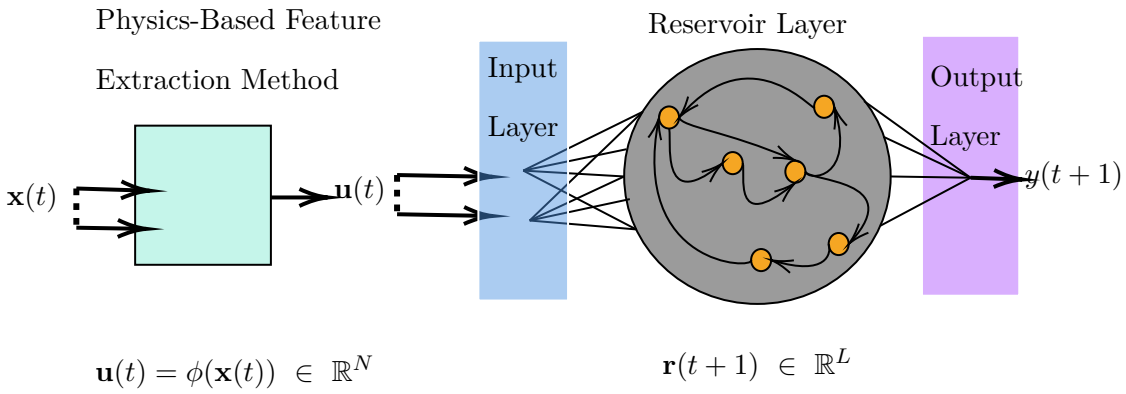


Figure 4.3: Reservoir Computing

The green block represents a feature extraction method. This is applied to the raw data such that the detection of medium intensity crashes becomes more feasible. We explain this further in the Results section. The first layer of RC consists of nodes to represent the time dependent input. Let the extracted input be denoted as $\mathbf{u}(t) \in \mathbb{R}^N$. The input is a vector of features with N dimension at time t. The input is connected to the second layer (called a reservoir). The connections between the input and the reservoir are represented by the matrix $\mathbf{W}_{in} \in \mathbb{R}^{L \times N}$. The input matrix elements are drawn from a uniform distribution in $[-\sigma, \sigma]$ [96]. The reservoir consists of L neurons as shown in Figure 4.3, where the topology of the reservoir is cyclic. The reservoir connections are represented by the matrix $\mathbf{W}_r \in \mathbb{R}^{L \times L}$. The state vector of the reservoir is denoted by $\mathbf{r}(t) \in \mathbb{R}^L$. The vector $\mathbf{r}(t)$ evolves over time

with a discrete time step. The reservoir $r(t)$ is updated according to equation (4.1).

$$\mathbf{r}(t+1) = (1-\alpha)\,\mathbf{r}(t) + \alpha\,\tanh\Big(\mathbf{W}_{in}\mathbf{u}(t) + \mathbf{W}_r\mathbf{r}(t)\Big) \tag{4.1}$$

The parameter $\alpha \in [0,1]$ determines how much to rely on the previous reservoir state and input. This parameter is called a leakage rate. That is to say, the reservoir evolves in a slow rate as α goes to zero. The matrix $\mathbf{W}_r$ is the adjacency matrix of the reservoir. This matrix is a sparse random Erdős – Rényi matrix. The elements of $\mathbf{W}_r$ are scaled such that $\rho(\mathbf{W}_r) < 1$, where ρ is the largest absolute eigenvalue. This is a necessary and sufficient condition for the echo state property of RC [60]. This property states that the effect of $\mathbf{u}(t)$ and $\mathbf{r}(t)$ on a future reservoir state $\mathbf{r}(t+\tau)$ vanishes gradually as $\tau \to \infty$. The function $\tanh(\mathbf{x})$ is defined as the hyperbolic tangent to the components of $\mathbf{x}$.

A traditional approach for interpreting the reservoir outputs for classification tasks is to classify the input based on the reservoir state $r(t)$ [109, 55], which is given by

$$y(t) = \mathbf{w}_{out}\mathbf{r}(t) + b, \tag{4.2}$$

where $y(t) \in \mathbb{R}$ is the output of the reservoir (being normal and abnormal), $\mathbf{w}_{out} \in \mathbb{R}^{1\times L}$ is the readout weights vector, and $b \in \mathbb{R}$ is a bias term. Both of these parameters are optimized during the training stage (i.e., for d trading) [86]. The readout weights and bias are determined such that $l(y(t), y_{target}(t))$ is minimized over a time period $t = [0, T-1]$ and d trading days. That is to say, if $u(t)$ belongs to a class 1, the weight vector and bias are computed such that $\mathbf{w}_{out}\mathbf{r}(t) + b \approx 1$. In the testing or prediction stage, when a new trading day is encountered with a new $\mathbf{u}(t)$, the system evolves according to (4.1) and the prediction of the class of the system at time t is obtained from (4.2) for $1 \le t \le T-1$.

4.4.3 Proposed Cost-Sensitive RC

Let the $\Xi \in \mathbb{R}^{2\times 2}$ denote the cost or loss matrix in the classification problem. The cost $\xi_{c_1 c_2}$ represents the cost of misclassifying an instance that belongs to c_1 (normal) into c_2 (abnormal). The diagonal of Ξ is usually set to zero. In the classification problem, a classifier aims to minimize the expected risk $\mathcal{R}(c_j \mid \mathbf{r})$, where c_j is the predicted class at time $t+1$. The expected risk is given by

$$\mathcal{R}(c_j \mid \mathbf{r}) = \sum_k \xi_{c_k c_j} P(c_k \mid \mathbf{r}), \text{ for } k \in \{1, 2\}, \tag{4.3}$$

where $P(c_k \mid \mathbf{r})$ is the posterior probability of a class c_k given an instance $\mathbf{r}$. The optimal class is the one that minimizes the Bayes risk

$$\hat{c}_j = \arg\min_{c_j} \mathcal{R}(c_j \mid \mathbf{r}) = \arg\min_{c_j} \mathbb{E}_{\mathcal{D}}[\Xi], \tag{4.4}$$

where $\mathcal{D}$ is the data in the form of input/output pairs.

The Bayes risk is usually computed empirically since the conditional distribution in (4.4) is not trivially determined. Suppose that we have d trading days with the data $\mathcal{D} = \{\mathbf{R}, y_{target}\}^d$, where $\mathbf{R} \in \mathbb{R}^{L\times[0,T-1]}$ is the matrix of the reservoir evolution vector with length L, covering a period between 0 and $T-1$, and $y_{target} \in \mathbb{R}^{[0,T-1]}$ is the target label vector for the time period $[0, T-1]$. In particular, $\mathbf{r}^{(i)}(t)$ denotes the reservoir vector $\mathbf{r}$ at time t and trading day i and $y^{(i)}_{target}(t) \in \{0, 1\}$ denotes the target label at time t and trading day i.

Given the imbalance data $\mathcal{D}$, the empirical risk for the training period (i.e., over a period $t \in [0, T-1]$ and d trading days) is defined as follows:

$$J(\mathbf{w}_{out}, b) = \hat{\mathcal{R}}_l = \frac{1}{T * d} \sum_{i=1}^{d} \sum_{t=0}^{T-1} l\Big(\Xi, y^{(i)}_{target}(t), y^{(i)}(t)\Big), \tag{4.5}$$

where $y^{(i)}(t)$ is the output of the RC at time t for the ith trading day, and $l(\cdot)$ is the loss

function, which is parameterized by Ξ . When the loss function is cost-insensitive, $\xi_{c_j,c_k} = 1$ when $c_j \neq c_k$, and zero otherwise. Due to the imbalance data distribution, we assume that $\xi_{c_j,c_k} \in \mathbb{R}^+$ when $c_i \neq c_j$. In this chapter, we use the Cross Entropy loss function. This can be easily derived as follows. Let the probability of the reservoir state belongs to class=1 (abnormal) be determined by the transformation of the output $y(t)$ such that it is between zero and one. That is, we apply a sigmoid function to the output node.

Equation (4.2) becomes

$$y(t) = \frac{1}{1 + e^{-\mathbf{w}_{out}\mathbf{r}(t)-b}}, \tag{4.6}$$

where $\mathbf{w}_{out} \in \mathbb{R}^{1\times L}$ is the readout weights vector, and $b \in \mathbb{R}$ is the bias. It should be noted that $y(t)$ is a function of $\mathbf{w}_{out}$ but that dependency is omitted in our notation.

The probabilities of a normal and abnormal instance at time t, given the reservoir state at time t, are defined as

$$P(y_{target}(t) = 1 \mid \mathbf{r}(t); \mathbf{w}_{out}) = y(t), \tag{4.7}$$

$$P(y_{target}(t) = 0 \mid \mathbf{r}(t); \mathbf{w}_{out}) = 1 - y(t). \tag{4.8}$$

The previous probabilities can be written in the compact form

$$P(y_{target}(t) \mid \mathbf{r}(t); \mathbf{w}_{out}) = (y(t))^{y_{target}(t)}(1 - y(t))^{1-y_{target}(t)}. \tag{4.9}$$

The Cross Entropy loss function is basically -1 times the *log* of the previous compact form. Thus, the empirical risk is given by

$$J(\mathbf{w}_{out}, b) = -\frac{1}{T * d}\sum_{i=1}^{d}\sum_{t=0}^{T-1}\Big[\Big(\xi_{c_2c_1}y^{(i)}_{target}(t)\Big)log\Big(y^{(i)}(t)\Big) + \xi_{c_1c_2}\Big(1 - y^{(i)}_{target}(t)\Big)log\Big(1 - y^{(i)}(t)\Big)\Big], \tag{4.10}$$

where $\xi_{c_2c_1}$ is the cost of misclassifying a class 1 (i.e., abnormal) as class 0 (i.e., normal). If $\xi_{c_2c_1} > 1$, the false negative is decreased, hence the Recall is increased. While, setting $\xi_{c_2c_1} < 1$ decreases the false positive and increases the Precision. In the previous cost function, we assumed that $\xi_{c_1c_2} = 1$.

If we are interested in the generalization of RC on a new data, we propose to use regularization. The regularization aims to reduce overfitting by reducing the variance while the bias is maintained. This method adds penalty to the cost function in equation (4.10) [37]. We consider adding a penalty term L_2 norm to the cost function

$$\begin{aligned} J(\mathbf{w}_{out}, b) = -\frac{1}{T * d}\sum_{i=1}^{d}\sum_{t=0}^{T-1}\Big[\Big(\xi_{c_2c_1}y^{(i)}_{target}(t)\Big)log\Big(y^{(i)}(t)\Big) + \Big(1 - y^{(i)}_{target}(t)\Big)log\Big(1 - y^{(i)}(t)\Big)\Big] \\ + \frac{\lambda}{2} \mid\mid \mathbf{w}_{out} \mid\mid_2^2, \end{aligned} \tag{4.11}$$

where $\lambda \geq 0$ is a regularization parameter. We refer to (4.11) as *Cost-Sensitive RC*.

The optimal parameters $\mathbf{w}_{out}$ and b can be computed by solving the optimization problem

$$\hat{\mathbf{w}}_{out}, \hat{b} = \arg\min_{\mathbf{w}_{out}, b}\Big(J(\mathbf{w}_{out}, b)\Big). \tag{4.12}$$

The classical approach to solve the optimization problem in (4.12) is to use a stochastic, first-order optimization algorithm such as Stochastic Gradient Decent (SGD). In this chapter, we use SGD to optimize the proposed *Cost-Sensitive RC*.

4.4.4 Establishing a Hybrid Model

In the previous section, we used only the past information about the markets to predict if the next class is normal or abnormal. This approach (hereafter called reservoir approach) makes little assumptions about the underlying mechanism (hereafter called physics-based approach) of equity markets. In this section, we propose a hybrid approach that combines the reservoir approach and the the physics-based approach to improve classification accuracy. Our approach is similar in spirit to that in [97], but here we consider the classification problem with a newly developed physics-based model called the MEM model.

In Chapter 3, we obtained predictions from the MEM model about the state of the markets $\mathbf{u}(t)$ for $t+1$ when knowledge about the markets is available up to time t. Let the one-step ahead prediction of the MEM model be given by

$$\hat{\mathbf{u}}(t+1) = \mathcal{U}[\mathbf{u}(t)], \tag{4.13}$$

where $\hat{\mathbf{u}}$ is the predicted state of the markets. The predictive model $\mathcal{U}$ is obtained using the proposed integrative algorithm in Chapter 3. The structure of the proposed hybrid approach is shown in Figure 4.4.

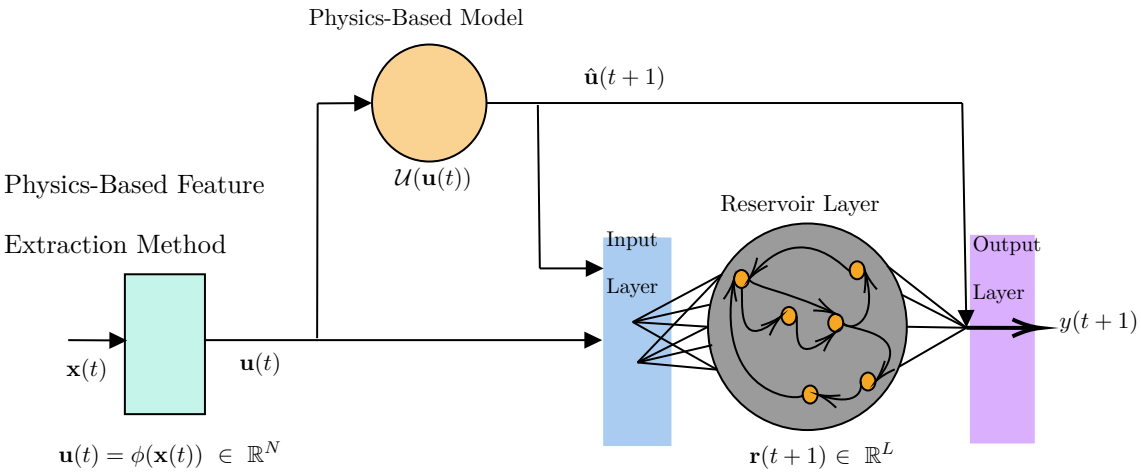


Figure 4.4: Hybrid RC

In the hybrid approach, the reservoir and output equations are functions of $\hat{\mathbf{u}}(t+1)$, which is the predicted input at $t+1$. These equations are given by

$$\mathbf{r}(t+1) = (1-\alpha)\,\mathbf{r}(t) + \alpha\,\tanh\Big(\mathbf{W}_{in}\Big(\mathbf{u}(t) \oplus \hat{\mathbf{u}}(t+1)\Big) + \mathbf{W}_r\mathbf{r}(t)\Big) \tag{4.14}$$

$$y(t+1) = f(\hat{\mathbf{u}}(t+1) \oplus \mathbf{r}(t+1); \mathbf{w}_{out}, b), \tag{4.15}$$

where $W_{in} \in \mathbb{R}^{L\times 2N}$ is the input matrix, f is the sigmoid function, and $\mathbf{w}_{out} \in \mathbb{R}^{1\times(N+L)}$ is the readout weights vector.

4.5 Results and Discussion

In this section, we apply RC to detect abnormalities in dataset from normal and abnormal trading days. We determine the accuracy of the proposed algorithm in detecting medium intensity crashes or abnormalities. These abnormalities occur during the trading day and have an overall affect on the markets. The data processing is detailed in the next section. In the performance measures, we explain how to test the performance of RC. Then, we show the implementation procedure of RC. The results section concludes with the sensitivity analysis of the model.

4.5.1 Data

The primary information used in this research is the raw data of stock prices traded in three major equity markets in the United States of America, namely: New York Stock Exchange (NYSE), NASDAQ, and American Stock Exchange (AMEX). On any trading day, approximately 4,000 shares were collected. Data frequency is one minute (i.e., the stock movement is recorded every minute). This frequency is very important for the detection of medium-frequency abnormalities. To begin, the data was cleaned to remove inactive stocks (a stock is active if it is traded more than 75% of a trading day). When a stock was not traded at a particular minute, we assumed that its price is similar to the last traded price.

This assumption is simple and known to researchers as it does not require any interpolations, which may affect the quality of the data.

4.5.2 Implementation

In this section, we discuss in detail the implementation of the proposed approaches. In both approaches, the reservoir input is not the raw stock price data. Our objective in this research is to detect medium intensity crashes. These crashes are macroscopic to most of the markets. Therefore, we employ the feature space that we established in the previous two chapters. During the training period, only data from $0 \leq t \leq T-1$ and d trading days are available. Such data includes the macroscopic features $\mathbf{u}(t)$. In the hybrid approach, additional information about the predictions is available as well, which is $\mathcal{U}[\mathbf{u}(t)]$.

4.5.2.1 Establishing the Feature Space ($\phi(\mathbf{x})$) for Detecting Medium Intensity Crashes

Following the pre-processing of data, time and space were diescretized. The time t was discretized into m equally spaced intervals. Each interval $\Delta t = 1$ matches the data frequency. Space, which is in \$, was discretized into n price ranges. The discretization size of space Δx was assumed to be one as well. As detailed in an earlier section, the set of features used in this research is a hybrid set of macroscopic variables, density and velocity, and information obtained from the MEM model. The density ρ and velocity v at a price x and time t were calculated as follows:

$$\rho(x,t) = \frac{N(x,t)}{\Delta x}, \tag{4.16}$$

$$v(x,t) = \frac{1}{N(x,t)} \sum_{j=1}^{N(x,t)} v_j(t), \tag{4.17}$$

where $N(x,t)$ represents the number of stocks in price x and time t.

After obtaining the macroscopic variables which represent the overall markets state, we include a second set of features. These features are composed of the model's parameters, which we found relevant to describe equity markets during high volatility and shocks periods. For details regarding the prediction methodology for both the parameters and state variables, we refer the reader to Chapter 3.

The macroscopic features are processed further to obtain y_{target} for each minute in the market. The 95% confidence interval at time t is determined from the previous ten minutes information. A feature at time t is labeled as abnormal if it is not within the confidence interval. While, a feature vector $\mathbf{u}(t)$ is labeled as abnormal at time t when at least five of its features are abnormal at that time.

4.5.2.2 Performance Measures

To have an adequate assessment of the classification results for the imbalanced data, we use traditional and other evaluation matrices that are suitable for imbalanced data. One of the most traditional methods is Accuracy. Consider the two-class classification problem presented in this research. Let {True Normal, True Abnormal} be the true normal and abnormal class labels and {Predicted Normal, Predicted Abnormal} be the predicted normal and abnormal class labels. The classification results can be summarized by a confusion matrix, as shown in Table 4.1. We assume that the majority class represents the negative class and the minority class represents the positive class.

		Predicted Class	
		Predicted Normal	Predicted Abnormal
True Class	True Normal	True Negative (TN)	False Positive (FP)
	True Abnormal	False Negative(FN)	True Positive (TP)

Table 4.1: Confusion Matrix

From the table, the Accuracy can be computed as

$$\text{Accuracy} = \frac{TP + TN}{TN + FP + FN + TP} \tag{4.18}$$

The accuracy provides a simple measure to determine the performance of the classification method on the datasets. This measure is not helpful for cases when data is imbalanced, as it is susceptible to changes in data distribution. In our datasets, the abnormal examples represent a small portion of the datasets so that a naive classifier might classify all examples as normal and still provide high accuracy. However, this metric doesn't take into the account that all abnormal examples are misclassified. Many researchers have studied the inadequacy of the accuracy measure for imbalanced data [124, 91, 99]. The central issue with this measure is that in the presence of imbalanced data, it becomes challenging to conduct a consistent analysis over different data types (i.e., balanced and imbalanced data) [58].

We complement our analysis by using other evaluation matrices to provide reliable analysis for imbalanced data including: Precision, Recall, F-measure, and Area Under the ROC Curve (AUC).

$$\text{Precision} = \frac{TP}{TP + FP}. \tag{4.19}$$

The Precision measures how accurate the model is in labeling examples as abnormal when they are abnormal. When the cost of false positive (i.e.,when the cost of alarming the market about a crash, when in fact, there is not one) is high, this measure should be high.

$$\text{Recall} = \frac{TP}{TP + FN}. \tag{4.20}$$

Recall, on the other hand, measures the accuracy of the model in detecting abnormal examples. If the cost of the false negative (i.e., when the model does not recognize a crash), is high the Recall measure should be high. The previous two measures are more effective when used together and correctly [58]. Another measure that combines the advantages of Recall and Precision is F-measure, which is defined as

$$\text{F-Measure} = \frac{2 * \text{Recall*Precision}}{\text{Recall+Precision}}. \tag{4.21}$$

This measure is preferable when the goal is to balance both the Precision and Recall. In this research, we report all the previously mentioned measures and discuss them accordingly.

4.5.2.3 Implementation Results

We should first discuss the training procedure in this research. We have 20 trading days. Each is about 391 minutes long. The trading days are divided into S segments of length T^*. We assume that each segment is independent from all other segments on the same day and other days (i.e., Markov property of equity markets). We further let each segment start at time 0 and end at time $T^* - 1$. Thus, the cost function in (4.11) is modified slightly such that d is replaced by $d \times S$ and $T - 1$ is replaced by $T^* - 1$. This way the proposed models can learn from a wider range of examples and patterns. In this stage, we train the reservoir and hybrid approaches on 75% of the $d \times S$ segments. 15% of the segments are used for testing. The extra segments are used for validation. The validation set is used to validate and tune the reservoir and regularization parameters, and misclassfiaction cost. These parameters include: $L, \sigma, \rho, \alpha, \lambda$, and $\xi_{c_2 c_1}$. In validation [47], we aim to identify the optimal set of parameters such that the Accuracy, Recall, and Precision are maximized, while placing more emphasis on Recall.

In this chapter, we only present the results of testing the proposed approaches on datasets with abnormalities. The class distribution of the testing datasets is shown in Figure 4.5. As can be seen from the figure, the class distribution is imbalanced. This is not surprising since medium intensity crashes do not happen often during trading days. As we explained earlier, such crashes vanish in a few minutes.

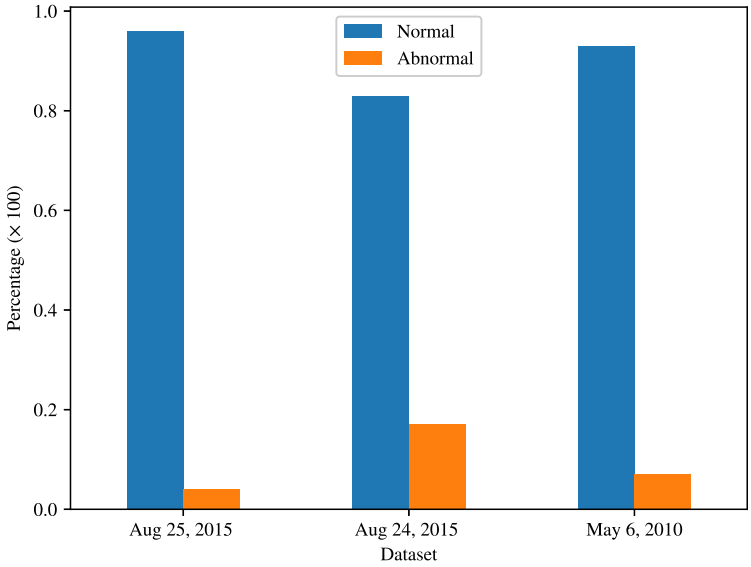


Figure 4.5: Class Distribution of Testing Datasets

The results of testing the reservoir and hybrid approaches are shown in Table 4.3. For the results in Table 4.3, we utilized the following set of parameters:

Parameter	**Value**
L	500
α	0.9
σ	1
ρ	1
λ	0.001
$\xi_{c_2c_1}$	20

Table 4.2: Set of Parameters

Hybrid Approach					Reservoir Approach				
Dataset #	**Precision**	**Recall**	**F-Measure**	**Accuracy**	**Dataset #**	**Precision**	**Recall**	**F-Measure**	**Accuracy**
Aug 25, 2015	1.0	1.0	1.0	1.0	**Aug 25, 2015**	0.11	0.50	0.18	0.70
Aug 24, 2015	0.67	0.80	0.74	0.90	**Aug 24, 2015**	0.50	0.60	0.55	0.83
May 6, 2010	1.0	1.0	1.0	1.0	**May 6, 2010**	0.25	1.0	0.40	0.90

Table 4.3: Classification Results on the Testing Datasets

It can be clearly seen from Table 4.3 that both approaches yield reasonable results. The hybrid approach seems to provide consistent results across different datasets. The Precision and Recall of the hybrid approach are high, while the Recall is acceptable for the reservoir approach. It should be noted that the reservoir parameters in both approaches were tuned such that the Recall is high even at the expense of the Precision. But, we can easily see that the Precision is high enough in most of the dataset. The F-measure for the hybrid approach is much higher than the reservoir approach, indicating the superior performance of the hybrid approach.

The classification task that we study is not trivial, as it involves making future predictions about the class of the input; not only detecting the current class. The hybrid approach proved to be a successful method. The reason behind this is that the hybrid approach includes additional information about the future from a physics-based model.

4.6 Conclusion

In this chapter, we tackled a classification problem in equity markets that has received little attention in the literature. With the spread of new trading paradigms in equity markets, a new form of crashes (defined in this research as medium intensity crashes) emerged. This problem is challenging and important. We presented a novel approach based on a neural network called RC. RC is a family of Recurrent Neural Networks in which some of the network connections are sparse, making it more efficient during training. We extended the

present work on RC to include a cost-sensitive parameter such that more weight can be added to the minority class. The introduced parameter was tuned to yield a high Recall performance measure. This is significant in the markets as undetected crashes (i.e., false negative) are undesirable. To improve the classification performance of RC, we proposed a hybrid model that combines the classical RC and a physics-based model. In Chapter 2 and 3, we introduced the MEM model. In this chapter, we utilized the MEM model as the physics-based model used in our hybrid approach. The hybrid approach has advantage over the classical approach as information about the underlying mechanisms of the markets are considered. Also, predictions are utilized in RC to improve the class predictions.

We trained and tested the proposed approaches on high frequency data. The obtained results indicate that the hybrid approach is promising in detecting and predicting markets anomalies. The results show that the Recall of the predicted class is high in most of the datasets. The results of the classical approach are considered satisfactory.

Chapter 5

Conclusion

One of the central challenges in equity markets is how to predict market abnormalities. This issue has been the focus of a considerable body of research and yet, efficient models for monitoring equity markets have been lacking. This work constitutes a major step forward in our understating of the markets. Other models have a major drawback in that they analyze markets microscopically while neglecting the overall interactions. To overcome this drawback, we introduced a macroscopic model to bring more clarity about markets' behavior under the assumption that stocks move as fluid particles. This analogy between fluid dynamics and stocks is original, allowing markets to be governed by physics-like principles and laws.

In the first part of this dissertation, we introduced the foundation and analogy between physics and equity markets. This analogy considers the overall market activities. We treated stocks as particles and macroscopic variables, such as density and velocity which were defined. To determine the evaluation of the macroscopic variables, we proposed a system of stochastic partial differential equations based on the conservation laws in fluid dynamics. The system was validated with data from the flash crash day, showing that adopting the macroscopic perspective is promising. The second part of this dissertation extended the previous part by examining the structural proprieties of the model. We theoretically analyzed the model and provided insights into the type of solutions it generates. Also, we proposed an efficient algorithm to solve the macroscopic model dynamically. The proposed algorithm was tested

on datasets with abnormalities and the results showed that abnormalities can be identified effectively. In the final part, we tackled a problem that, to our knowledge, has not been addressed in the literature. Specifically, this is the problem of detecting medium intensity crashes, which are crashes that occur in a given trading day. The scarcity of these crashes creates imbalance in the data. To overcome this challenge, we proposed a cost-sensitive model based on RC. We also extended the classical RC to process information from a physics-based model. The results of implementing the proposed hybrid approach showed high accuracy in detecting medium intensity crashes.

While this research has successfully developed monitoring tools, it could be extended in interesting ways. For instance, the model presented in Chapter 2 can be extended to include other financial markets, such as derivatives and foreign exchange. The extended model might offer a macroscopic look at financial markets by connecting them together and allowing the information to flow from one market to another. Extensions to Chapter 3 include employing higher resolution numerical methods when solving the deterministic homogeneous and inhomogeneous MEM models. Another potential extension is to utilize other filtering techniques, such as particle and sigma-points, to filter and estimate the stochastic inhomogeneous MEM model. These extensions are important for providing more accurate predictions of macroscopic variables. In Chapter 4, we proposed a hybrid approach based on RC. This approach is not limited to RC, but can also be used with other machine learning techniques.

Bibliography

Bibliography

[1] A. AlShelahi, N. Fougere, T. Zhang, and R. Saigal. Macroscopic look at the Equity Markets. Paper presented at the Financial Stability Conference, Ann Arbor, October 22-23, 2015.

[2] A. AlShelahi and R. Saigal. Insights into the macroscopic behavior of equity markets: Theory and application. *Physica A: Statistical Mechanics and its Applications*, 505:778–793, 2018.

[3] W. Aswolinskiy, R. F. Reinhart, and J. Steil. Time series classification in reservoir- and model-space: a comparison. In *IAPR Workshop on Artificial Neural Networks in Pattern Recognition*, pages 197–208. Springer, 2016.

[4] L. Bachelier. *Théorie de la spéculation*. Gauthier-Villars, 1900.

[5] A. Bagnall, J. Lines, A. Bostrom, J. Large, and E. Keogh. The great time series classification bake off: a review and experimental evaluation of recent algorithmic advances. *Data Mining and Knowledge Discovery*, 31(3):606–660, 2017.

[6] P. Bak. *How nature Works: The Science of Self-Organized Criticality*. Springer Science & Business Media, 2013.

[7] R. J. Barro and J. F. Ursúa. Stock-market crashes and depressions. Technical report, National Bureau of Economic Research, 2009.

[8] A. Belgrano and J. H. Brown. Ecology: Oceans under the macroscope. *Nature*, 419(6903):128–129, 2002.

[9] Y. Bengio, P. Simard, and P. Frasconi. Learning long-term dependencies with gradient descent is difficult. *IEEE transactions on neural networks*, 5(2):157–166, 1994.

[10] F. M. Bianchi, E. De Santis, A. Rizzi, and A. Sadeghian. Short-term electric load forecasting using echo state networks and pca decomposition. *Ieee Access*, 3:1931–1943, 2015.

[11] C. H. Bishop, B. J. Etherton, and S. J. Majumdar. Adaptive sampling with the ensemble transform kalman filter. part i: Theoretical aspects. *Monthly weather review*, 129(3):420–436, 2001.

[12] Bloomberg L.P. Flash crash trader Sarao pleads guilty to fraud, spoofing, 2016.

[13] J.-P. Bouchaud. The endogenous dynamics of markets: price impact and feedback loops. *arXiv preprint arXiv:1009.2928*, 2010.

[14] J.-P. Bouchaud and M. Mézard. Wealth condensation in a simple model of economy. *Physica A: Statistical Mechanics and Its Applications*, 282(3):536–545, 2000.

[15] R. Brown. A brief account of microscopical observations made in the months of June, July and August 1827, on the particles contained in the pollen of plants; and on the general existence of active molecules in organic and inorganic bodies. *Philosophical Magazine Series 2*, 4(21):161–173, 1828.

[16] G. Burgers, P. Jan van Leeuwen, and G. Evensen. Analysis scheme in the ensemble kalman filter. *Monthly weather review*, 126(6):1719–1724, 1998.

[17] Businessinsider. Amazon stock flash crashes and then quickly recovers, 2017.

[18] J. C. Butcher. *Numerical methods for ordinary differential equations*. John Wiley & Sons, 2016.

[19] P. Buteneers, D. Verstraeten, B. Van Nieuwenhuyse, D. Stroobandt, R. Raedt, K. Vonck, P. Boon, and B. Schrauwen. Real-time detection of epileptic seizures in animal models using reservoir computing. *Epilepsy research*, 103(2-3):124–134, 2013.

[20] E. P. Carden and J. M. Brownjohn. Arma modelled time-series classification for structural health monitoring of civil infrastructure. *Mechanical systems and signal processing*, 22(2):295–314, 2008.

[21] A. Chakraborti and B. K. Chakrabarti. Statistical mechanics of money: How saving propensity affects its distribution. *The European Physical Journal B-Condensed Matter and Complex Systems*, 17(1):167–170, 2000.

[22] A. Chatterjee, B. K. Chakrabarti, and S. Manna. Pareto law in a kinetic model of market with random saving propensity. *Physica A: Statistical Mechanics and Its Applications*, 335(1):155–163, 2004.

[23] N. V. Chawla, N. Japkowicz, and A. Kotcz. Special issue on learning from imbalanced data sets. *ACM Sigkdd Explorations Newsletter*, 6(1):1–6, 2004.

[24] H. Chen, F. Tang, P. Tino, and X. Yao. Model-based kernel for efficient time series analysis. In *Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 392–400. ACM, 2013.

[25] T. Choudhry, F. I. Papadimitriou, and S. Shabi. Stock market volatility and business cycle: Evidence from linear and nonlinear causality tests. *Journal of Banking & Finance*, 66:89–101, 2016.

[26] K.-C. Chu, L. Yang, R. Saigal, and K. Saitou. Validation of stochastic traffic flow model with microscopic traffic simulation. In *Automation Science and Engineering (CASE), 2011 IEEE Conference on Automation Science and Engineering*, pages 672–677. IEEE, 2011.

[27] D. Cliff, R. Calinescu, J. Keen, T. Kelly, M. Kwiatkowska, J. McDermid, R. Paige, and I. Sommerville. *The UK Large-Scale Complex IT Systems (LSCITS) Initiative*, 2010. manuscript available from http://lscits. cs. bris. ac. uk/docs/LSCITSoverview2010. pdf.

[28] S. Cordier, L. Pareschi, and C. Piatecki. Mesoscopic modeling of financial markets. *Journal of Statistical Physics*, 134(1):161–184, 2009.

[29] S. Cordier, L. Pareschi, and G. Toscani. On a kinetic model for a simple market economy. *Journal of Statistical Physics*, 120(1-2):253–277, 2005.

[30] R. Courant, K. Friedrichs, and H. Lewy. Über die partiellen differenzengleichungen der mathematischen physik. *Mathematische annalen*, 100(1):32–74, 1928.

[31] R. De Luca, M. Di Mauro, A. Falzarano, and A. Naddeo. A hydrodynamic model for cooperating solidary countries. *The European Physical Journal B*, 90(7):134, 2017.

[32] A. Deihimi and H. Showkati. Application of echo state networks in short-term electric load forecasting. *Energy*, 39(1):327–340, 2012.

[33] J. Döpke, U. Fritsche, and C. Pierdzioch. Predicting recessions with boosted regression trees. *International Journal of Forecasting*, 33(4):745–759, 2017.

[34] K. Doya. Bifurcations in the learning of recurrent neural networks. In *Circuits and Systems, 1992. ISCAS'92. Proceedings., 1992 IEEE International Symposium on*, volume 6, pages 2777–2780. IEEE, 1992.

[35] A. Dragulescu and V. M. Yakovenko. Statistical mechanics of money. *The European Physical Journal B-Condensed Matter and Complex Systems*, 17(4):723–729, 2000.

[36] B. Düring and G. Toscani. Hydrodynamics from kinetic models of conservative economies. *Physica A: Statistical Mechanics and its Applications*, 384(2):493–506, 2007.

[37] X. Dutoit, B. Schrauwen, J. Van Campenhout, D. Stroobandt, H. Van Brussel, and M. Nuttin. Pruning and regularization in reservoir computing. *Neurocomputing*, 72(7-9):1534–1546, 2009.

[38] D. Easley, M. M. L. De Prado, and M. O'Hara. The microstructure of the" flash crash": Flow toxicity, liquidity crashes, and the probability of informed trading. *Journal of Portfolio Management*, 37(2):118, 2011.

[39] A. Ermogenous. Brownian Motion and its Applications in the Stock Market. In *Undergraduate Mathematics Day, Electronic Proceedings*. University of Dayton eCommons, 2006.

[40] L. Evans. *Partial Differential Equations (Graduate Studies in Mathematics vol 19)(Providence, RI: American Mathematical Society)*. Oxford University Press, 1998.

[41] L. C. Evans. *Partial Differential Equations*. American Mathematical Society, 2010.

[42] G. Evensen. Sequential data assimilation with a nonlinear quasi-geostrophic model using monte carlo methods to forecast error statistics. *Journal of Geophysical Research: Oceans*, 99(C5):10143–10162, 1994.

[43] D. Faranda, F. M. E. Pons, E. Giachino, S. Vaienti, and B. Dubrulle. Early warnings indicators of financial crises via auto regressive moving average models. *Communications in Nonlinear Science and Numerical Simulation*, 29(1-3):233–239, 2015.

[44] R. E. Farmer. The stock market crash of 2008 caused the great recession: Theory and evidence. *Journal of Economic Dynamics and Control*, 36(5):693–707, 2012.

[45] V. Filimonov and D. Sornette. Quantifying reflexivity in financial markets: Toward a prediction of flash crashes. *Physical Review E*, 85(5):056108, 2012.

[46] J. Fox and A. Sklar. *The myth of the rational market: A history of risk, reward, and delusion on Wall Street*. Harper Business New York, 2009.

[47] J. Friedman, T. Hastie, and R. Tibshirani. *The elements of statistical learning*, volume 1. Springer series in statistics New York, NY, USA:, 2001.

[48] K.-i. Funahashi and Y. Nakamura. Approximation of dynamical systems by continuous time recurrent neural networks. *Neural networks*, 6(6):801–806, 1993.

[49] A. Gelb. *Applied optimal estimation*. MIT press, 1974.

[50] F. A. Gers, J. Schmidhuber, and F. Cummins. Learning to forget: Continual prediction with lstm. *Neural Computation*, 12(10):2451–2471, 2000.

[51] S. K. Godunov. A difference method for numerical calculation of discontinuous solutions of the equations of hydrodynamics. *Matematicheskii Sbornik*, 89(3):271–306, 1959.

[52] A. Golub, J. Keane, and S.-H. Poon. High frequency trading and mini flash crashes. *arXiv preprint arXiv:1211.6667*, 2012.

[53] I. Goodfellow, Y. Bengio, A. Courville, and Y. Bengio. *Deep learning*, volume 1. MIT press Cambridge, 2016.

[54] A. Graves and J. Schmidhuber. Framewise phoneme classification with bidirectional lstm and other neural network architectures. *Neural Networks*, 18(5-6):602–610, 2005.

[55] L. Grigoryeva, J. Henriques, L. Larger, and J.-P. Ortega. Optimal nonlinear information processing capacity in delay-based reservoir computers. *Scientific reports*, 5:12858, 2015.

[56] M. Grinfeld, H. Lamba, and R. Cross. A mesoscopic stock market model with hysteretic agents. *Discrete Contin. Dyn. Syst., Ser. B*, 18:403–415, 2013.

[57] S. Haykin. *Kalman filtering and neural networks*, volume 47. John Wiley & Sons, 2004.

[58] H. He and E. A. Garcia. Learning from imbalanced data. *IEEE Transactions on Knowledge & Data Engineering*, 9:1263–1284, 2008.

[59] D. P. Hunt and D. Parry. Using echo state networks to classify unscripted, real-world punctual activity. In *Engineering Applications of Neural Networks*, pages 369–378. Springer, 2015.

[60] H. Jaeger. The âĂIJecho stateâĂİ approach to analysing and training recurrent neural networks-with an erratum note. *Bonn, Germany: German National Research Center for Information Technology GMD Technical Report*, 148(34):13, 2001.

[61] P. K. Jain. Financial market design and the equity premium: Electronic versus floor trading. *The Journal of Finance*, 60(6):2955–2985, 2005.

[62] N. N. Janenko. *The method of fractional steps*. Springer, 1971.

[63] C. M. Jarque and A. K. Bera. A test for normality of observations and regression residuals. *International Statistical Review/Revue Internationale de Statistique*, pages 163–172, 1987.

[64] N. Johnson, G. Zhao, E. Hunsader, H. Qi, N. Johnson, J. Meng, and B. Tivnan. Abrupt rise of new machine ecology beyond human response time. *Scientific reports*, 3:2627, 2013.

[65] N. F. Johnson, P. Jefferies, P. M. Hui, et al. Financial market complexity. *OUP Catalogue*, 2003.

[66] A. Joulin, A. Lefevre, D. Grunberg, and J.-P. Bouchaud. Stock price jumps: news and volume play a minor role. *arXiv preprint arXiv:0803.1769*, 2008.

[67] S. J. Julier and J. K. Uhlmann. New extension of the kalman filter to nonlinear systems. In *Signal processing, sensor fusion, and target recognition VI*, volume 3068, pages 182–194. International Society for Optics and Photonics, 1997.

[68] R. E. Kalman. A new approach to linear filtering and prediction problems. *Transactions of the ASME–Journal of Basic Engineering*, 82(Series D):35–45, 1960.

[69] A. Kampouraki, G. Manis, and C. Nikou. Heartbeat time series classification with support vector machines. *IEEE Trans. Information Technology in Biomedicine*, 13(4):512–518, 2009.

[70] M. Katzfuss, J. R. Stroud, and C. K. Wikle. Understanding the ensemble kalman filter. *The American Statistician*, 70(4):350–357, 2016.

[71] S. H. Khan, M. Hayat, M. Bennamoun, F. A. Sohel, and R. Togneri. Cost-sensitive learning of deep feature representations from imbalanced data. *IEEE transactions on neural networks and learning systems*, 29(8):3573–3587, 2018.

[72] A. Kirilenko, A. S. Kyle, M. Samadi, and T. Tuzun. The flash crash: High-frequency trading in an electronic market. *The Journal of Finance*, 2017.

[73] A. Kirman. The economic crisis is a crisis for economic theory. *CESifo Economic Studies*, 56(4):498–535, 2010.

[74] A. W. Kleidon. Stock market crashes. *Handbooks in Operations Research and Management Science*, 9:465–495, 1995.

[75] P. Krugman. How did economists get it so wrong? *New York Times*, 2(9):2009, 2009.

[76] B. L. Kyer and G. E. Maggs. Macroeconomic hydraulics reconsidered: Teaching leakages and injections from the classical perspective. *Journal of Economics and Finance Education*, 10(1):14–18, 2011.

[77] W. G. Lawson and J. A. Hansen. Implications of stochastic and deterministic filters as ensemble-based data assimilation methods in varying regimes of error growth. *Monthly weather review*, 132(8):1966–1981, 2004.

[78] P. D. Lax. *Hyperbolic systems of conservation laws and the mathematical theory of shock waves*, volume 11. SIAM, 1973.

[79] J. Lei, P. Bickel, and C. Snyder. Comparison of ensemble kalman filters under non-gaussianity. *Monthly Weather Review*, 138(4):1293–1306, 2010.

[80] R. J. LeVeque. Conservative methods for nonlinear problems. In *Numerical Methods for Conservation Laws*, pages 122–135. Springer, 1990.

[81] R. J. LeVeque. *Numerical methods for conservation laws*, volume 132. Springer, 1992.

[82] R. J. LeVeque. *Finite volume methods for hyperbolic problems*, volume 31. Cambridge university press, 2002.

[83] M. Levy and S. Solomon. New evidence for the power-law distribution of wealth. *Physica A: Statistical Mechanics and Its Applications*, 242(1–2):90–94, 1997.

[84] O. Linton and S. Mahmoodzadeh. Implications of high-frequency trading for security markets. *Annual Review of Economics*, 0, 2018.

[85] A. W. Lo. The adaptive markets hypothesis: Market efficiency from an evolutionary perspective. 2004.

[86] Z. Lu, J. Pathak, B. Hunt, M. Girvan, R. Brockett, and E. Ott. Reservoir observers: Model-free inference of unmeasured variables in chaotic systems. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 27(4):041102, 2017.

[87] M. Lukoševičius and H. Jaeger. Reservoir computing approaches to recurrent neural network training. *Computer Science Review*, 3(3):127–149, 2009.

[88] T. Lux. Applications of statistical physics in finance and economics. Technical report, Kieler Arbeitspapiere, 2008.

[89] Q. Ma, L. Shen, W. Chen, J. Wang, J. Wei, and Z. Yu. Functional echo state network for time series classification. *Information Sciences*, 373:1–20, 2016.

[90] W. Maass, T. Natschläger, and H. Markram. Real-time computing without stable states: A new framework for neural computation based on perturbations. *Neural computation*, 14(11):2531–2560, 2002.

[91] M. A. Maloof. Learning when data sets are imbalanced and when costs are unequal and unknown. In *ICML-2003 workshop on learning from imbalanced data sets II*, volume 2, pages 2–1, 2003.

[92] R. N. Mantegna and H. E. Stanley. *Introduction to econophysics: Correlations and complexity in finance*. Cambridge University Press, 1999.

[93] E. W. Ngai, Y. Hu, Y. Wong, Y. Chen, and X. Sun. The application of data mining techniques in financial fraud detection: A classification framework and an academic review of literature. *Decision support systems*, 50(3):559–569, 2011.

[94] T. N. Nicholas. *The Black Swan: The impact of the highly improbable*. NY : Random, 2007.

[95] R. S. Palais and R. A. Palais. *Differential equations, mechanics, and computation*, volume 51. American Mathematical Soc., 2009.

[96] J. Pathak, B. Hunt, M. Girvan, Z. Lu, and E. Ott. Model-free prediction of large spatiotemporally chaotic systems from data: a reservoir computing approach. *Physical review letters*, 120(2):024102, 2018.

[97] J. Pathak, A. Wikner, R. Fussell, S. Chandra, B. R. Hunt, M. Girvan, and E. Ott. Hybrid forecasting of chaotic processes: using machine learning in conjunction with a knowledge-based model. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 28(4):041101, 2018.

[98] A. Prater. Spatiotemporal signal classification via principal components of reservoir states. *Neural Networks*, 91:66–75, 2017.

[99] F. J. Provost, T. Fawcett, et al. Analysis and visualization of classifier performance: Comparison under imprecise class and cost distributions. In *KDD*, volume 97, pages 43–48, 1997.

[100] Reuters. Minute by minute, Nasdaq chaos engulfed Facebook IPO, 2012.

[101] P. Richmond, J. Mimkes, and S. Hutzler. *Econophysics and physical economics*. Oxford University Press, 2013.

[102] A. Rodan, A. F. Sheta, and H. Faris. Bidirectional reservoir networks trained using svm

+

+ privileged information for manufacturing process modeling. *Soft Computing*, 21(22):6811–6824, 2017.

[103] R. Saigal and C. Chu. A dynamic stochastic travel time simulation model, 2008. University of Michigan.

[104] S. Scardapane and D. Wang. Randomness in neural networks: an overview. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 7(2):e1200, 2017.

[105] G. W. Schwert. Why does stock market volatility change over time? *The journal of finance*, 44(5):1115–1153, 1989.

[106] G. W. Schwert. Anomalies and market efficiency. *Handbook of the Economics of Finance*, 1:939–974, 2003.

[107] W. Shi, P. Shang, and A. Lin. The coupling analysis of stock market indices based on cross-permutation entropy. *Nonlinear Dynamics*, 79(4):2439–2447, 2015.

[108] Z. Shi and M. Han. Support vector echo-state machine for chaotic time-series prediction. *IEEE Trans. Neural Networks*, 18(2):359–372, 2007.

[109] M. C. Soriano, D. Brunner, M. Escalona-Morán, C. R. Mirasso, and I. Fischer. Minimal approach to neuro-inspired information processing. *Frontiers in computational neuroscience*, 9:68, 2015.

[110] D. Sornette. *Why stock markets crash: critical events in complex financial systems.* Princeton University Press, 2017.

[111] D. Sornette and G. Ouillon. Dragon-kings: Mechanisms, statistical methods and empirical evidence. *The European Physical Journal-Special Topics*, 205(1):1–26, 2012.

[112] D. Sornette and S. Von der Becke. Crashes and High Frequency Trading: An evaluation of risks posed by high speed algorithmic trading. *Foresight, the Future of Fomputerised Trading in Financial Markets, Driver Review*, 7:103–106, 2011.

[113] Y. Sun, M. S. Kamel, A. K. Wong, and Y. Wang. Cost-sensitive boosting for classification of imbalanced data. *Pattern Recognition*, 40(12):3358–3378, 2007.

[114] P. Tanisaro and G. Heidemann. Time series classification using time warping invariant echo state networks. In *Machine Learning and Applications (ICMLA), 2016 15th IEEE International Conference on*, pages 831–836. IEEE, 2016.

[115] M. K. Tippett, J. L. Anderson, C. H. Bishop, T. M. Hamill, and J. S. Whitaker. Ensemble square root filters. *Monthly Weather Review*, 131(7):1485–1490, 2003.

[116] E. F. Toro. *Riemann solvers and numerical methods for fluid dynamics: a practical introduction*. Springer Science & Business Media, 2013.

[117] E. Trentin, S. Scherer, and F. Schwenker. Emotion recognition from speech signals via a probabilistic echo-state network. *Pattern Recognition Letters*, 66:4–12, 2015.

[118] B. Vašíček, D. Žigraiová, M. Hoeberichts, R. Vermeulen, K. Šmídková, and J. de Haan. Leading indicators of financial stress: New evidence. *Journal of Financial Stability*, 28:240–257, 2017.

[119] J. Voit. From Brownian motion to operational risk: Statistical physics and financial markets. *Physica A: Statistical Mechanics and its Applications*, 321(1):286–299, 2003.

[120] J. B. Walsh. An introduction to stochastic partial differential equations. In *École d'Été de Probabilités de Saint Flour XIV-1984*, pages 265–439. Springer, 1986.

[121] E. A. Wan and R. Van Der Merwe. The unscented kalman filter for nonlinear estimation. In *Adaptive Systems for Signal Processing, Communications, and Control Symposium 2000. AS-SPCC. The IEEE 2000*, pages 153–158. Ieee, 2000.

[122] L.-X. Wang. Dynamical models of stock prices based on technical trading rules part i: the models. *IEEE Transactions on Fuzzy Systems*, 23(4):787–801, 2015.

[123] K. Watanabe, H. Takayasu, and M. Takayasu. Random walker in temporally deforming higher-order potential forces observed in a financial crisis. *Physical Review E*, 80(5):056110, 2009.

[124] G. M. Weiss. Mining with rarity: a unifying framework. *ACM Sigkdd Explorations Newsletter*, 6(1):7–19, 2004.

[125] T. R. Werner, T. Gubiec, R. Kutner, and D. Sornette. Modeling of super-extreme events: An application to the hierarchical Weierstrass-Mandelbrot continuous-time random walk. *The European Physical Journal-Special Topics*, 205(1):27–52, 2012.

[126] L. Yang. *Stochastic Traffic Flow Modeling and Optimal Congestion Pricing*. PhD thesis, University of Michigan, 2012.

[127] L. Yang, R. Saigal, C. Chu, and Y. Wan. Stochastic model for traffic flow prediction and its validation. In *Transportation Research Board 90th Annual Meeting*, 2011.

[128] Y. Yura, H. Takayasu, D. Sornette, and M. Takayasu. Financial brownian particle in the layered order-book fluid and fluctuation-dissipation relations. *Physical review letters*, 112(9):098703, 2014.

[129] W. Zhang and J. Wang. Nonlinear stochastic exclusion financial dynamics modeling and complexity behaviors. *Nonlinear Dynamics*, 88(2):921–935, 2017.

[130] Z.-H. Zhou and X.-Y. Liu. Training cost-sensitive neural networks with methods addressing the class imbalance problem. *IEEE Transactions on Knowledge and Data Engineering*, 18(1):63–77, 2006.