8-2013

Mixture of Poisson distributions to model discrete stock price changes.

Rasitha Rangani Jayasekare Kodippuli Thanthillage Dona

MIXTURE OF POISSON DISTRIBUTIONS TO MODEL DISCRETE STOCK
PRICE CHANGES

By

Rasitha Rangani Jayasekare Kodippuli Thanthillage Dona
B.Sc., Rajarata University of Sri Lanka, 2004
M.Sc., University of Sri Jayawardenepura, Sri Lanka, 2008
M.A., University of Louisville, Kentucky, USA, 2011

A Dissertation
Submitted to the Faculty of the
College of Arts and Sciences of the University of Louisville
in Partial Fulfillment of the Requirements
for the Degree of

Doctor of Philosophy

Department of Mathematics
University of Louisville
Louisville, KY

August 2013

MIXTURE OF POISSON DISTRIBUTIONS TO MODEL DISCRETE STOCK
PRICE CHANGES

Submitted by

Rasitha Rangani Jayasekare Kodippuli Thanthillage Dona

A Dissertation Approved on

July 10, 2013

(Date)

by the Following Reading and Examination Committee:

Dr. Ryan Gill (Dissertation Chair)

Dr. Kiseop Lee (Co-Chair)

Dr. Thomas Riedel

Dr. Robert Powers

Dr. Mehmed Kantardzic

DEDICATION

This dissertation is dedicated to my parents

Sarath Wickramasinghe and Nanda Wickramasinghe

who have given me invaluable educational opprotunities

and

my daughter Hasara Sethsandi.

ACKNOWLEDGEMENTS

ABSTRACT

MIXTURE OF POISSON DISTRIBUTIONS TO MODEL DISCRETE STOCK PRICE CHANGES

Rasitha Rangani Jayasekare Kodippuli Thanthillage Dona

July 10, 2013

An application of a mixture of Poisson distributions is proposed to model the discrete changes in stock price based on the minimum price movement known as 'tick-size'. The parameters are estimated using the Expectation-Maximization (EM) algorithm with a constant mixing probability as well as mixing probabilities which depend on order size. The model is evaluated using simulations and real data. Both the simulated and real data show reasonable estimates.

Several adjustments are made to the model implementation to improve the efficiency with user written codes for the Newton Raphson algorithm and also implementing one of the most recent versions of the EM algorithm (PEM). Both the improvements show an exponentially increasing efficiency to the implementation.

Further a Clustered Signed model is proposed to use summarized data to reduce the amount of data to be used in the model implementation using the discrete order sizes and the signs of the discrete stock price changes. The clustered model provided a significant time efficiency. A parametric bootstrap procedure is also considered to assess the significance of the order size on the mixing probabilities. The results show that the use of a variable mixture probability, which depends on the order size, is more appropriate for the model. The methods are illustrated with data from simulations and real data from Federal Express.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1
INTRODUCTION

1.1 Stock Markets

Stock market provides a way of fulfilling the needs of raising the capital of a company, where the ownership of a company is shared among several personnel via 'shares' or also known as 'stocks' (*Stock Market*, 2013). The stock market is divided into two sectors as primary market and secondary market. When the stock is first issues from a company the trading takes place in the primary market, followed by trading in the secondary market (*Stock Market*, 2013).

The London stock exchange is assumed to be the first stock market in the world which started in the early 16th century (*Stock Market History*, 2005). In 1792 the New York Stock Exchange (NYSE) started with just two dozen stock brokers in the city of New York as the first stock market in the United States of America (*Stock Market History*, 2005). Since then stock markets have been introduced and grown rapidly and have become one of the major financial topics in the daily life.

While stocks are traded constantly, it is interesting to know the factors that govern the stock market. This interesting question has been intensely studied for a long period of time and has led to many contributions in different fields. Different research has looked at the changes in stock price as a problem of 'detection' as well as a problem of 'prediction'.

1.1.1 Applications on Stock Market

The literature provides a large number of publications on stock market data using many data mining methodologies. Time series analysis, neural networks, regression models, support vector machines, liquidity effect model, markov chains and hybrid models are few of them.

Gavrishchaka and Banerjee (2006), present an application of Support Vector Machines (SVM) to extract information from highly dimensional and multi scaled stock market data for forecasting stock market volatility. Support vector machine is a supervised learning model that is used for regression analysis and data classification. SVM is well known for both linear and non-linear classification data. Huang *et al.* (2005) also uses support vector machines to forecast stock market movement directions. According to the authors, the model based on the SVM outperforms the other classification models in their research. Yu *et al.* (2009) also use an application of support vector machines to forecast stock market trends with an 'evolving least squares support vector machine'.

Liquidity Effect Model also has been used in stock data analysis. Liquidity Effect Model describes how the purchases and the sales of stocks in the stock market affect the changes of the stock price. Significant changes to the stock price and the locations of those significant changes can also be identified using liquidity effect. Gill *et al.* (2007) presents 'Computation of estimates in segmented regression and a liquidity effect model' on stock transactions data. A weighted least squares estimation along with a liquidity effect model is performed as a generalization to the liquidity effect model proposed by Cetin *et al* (2006).

Li and Liu (2009) presents an 'Application Study of Back-Propagation Neural Network on Stock Market Prediction' for efficient short term prediction of stock market based on a 3-layered feed forward neural network. Neural Networks are un-

supervised learning techniques in data mining. According to Zhou and Jie (2010) Back Propagation is one of the most often used neural network models which is based on a forward multi layered network which is trained by minimum mean square error. Zhou and Jie (2010) perform a stock market analysis based on a back propagation neural network. The authors research on stock transactions trend of price to predict the future trend of the stock market and the changes in the stock price based on the data from Chinese stock market. Enke and Thawornwong (2005) also use neural network for forecasting stock market returns. The research discusses the use of neural networks to uncover relationships of the stock market returns using data mining. Oh and Kim (2002) also uses a back propagation neural network to forecast change points to predict stock price index using a stock trading model based on chaotic analysis and a piecewise nonlinear model.

Time series analysis is another data mining technique that is widely used in stock market analysis. Time series analysis mines data recorded in a temporal sequence for both detection and predication of data. A 'Time Series Clustering Based on Independent Component Analysis' is presented by Guo *et al.* (2008). The methods employ a feature based approach to time series clustering that includes an independent component analysis and a modified k-means algorithm to overcome the difficulty of the time series clustering on stock data. Yaho and Kong (2008) presents an 'Application of Stream Data Time Series Pattern Reliance Mining in Stock Market Analysis' which uses stream data, static databases along with data mining. 'Evolutionary Time Series Segmentation for Stock Data Mining' (Chung *et al.* 2002) presents a transformation of stock transaction data into meaningful symbols like technical patterns to overcome the difficulty of the nature of multiple time series of stock data.

Fuzzy time series models provide advantages when clearly separated states are not possible and also when the linguistic values are available. Chen *et al.* (2007)

present a forecasting on stock price data based on a fuzzy time series model which employs a Fibonacci sequence for stock price forecasting. Weighted fuzzy time series models were used in stock index forecasting by Yu (2005). Jilani and Burney (2008) present a time-variant fuzzy time series forecasting model for forecasting stock prices. The proposed method uses a heuristic approach to define frequency-density-based partitions. The implementation of the model in forecasting of stock price in Taiwan stock exchange showed a higher accuracy compared to similar models in forecasting.

Due to an enormous interest in stock market, a large amount of research was performed with many different models and their variations and combinations. Huang and Jane (2009) researched on a hybrid models based on ARX, grey system and Rough Set theories to forecast stock market returns and portfolio selections. The authors present that hybrid models perform better than stand alone models and also produce greater results in forecasting.

Diaz et al. (2011) have applied a collection of knowledge discover techniques on stock intra-day trade prices to identify stock price manipulations. The techniques include regression, frequency outliers analysis, unsupervised learning techniques and supervised learning such as QUEST and C5.0. Greenwood and Thesmar (2011) predict volatility of the stock price based on the fragility by identifying the relationship between financial assets and non fundamental risks.

'Mixture of Compound Poisson Processes as Models tick-by-tick Financial Data' (Scalas *et al.* 2007) implements the idea of 'continuous random walk' and normal compound poisson processes. Plerou *et al.* (2002) performs an analysis of stock data to investigate on how the stock prices respond to changes in demand by identifying the relationship between stock price changes over time intervals base on a spin system. Podobnik *et al.* (2009) research on volume growth rate of a trade using a detrended cross-correlation analysis. The research further analyses

4

the properties of the volume change of trades and their relationship to stock price changes. Atsalakis and Valavanis (2009) present a survey of more than hundreds of research publications on stock market forecasting based on neural networks and fuzzy logics. The authors state that the successful forecasting is achieved when the minimum amount of input data are used to obtain the best results.

Behavioral economics and social media are also used in stock market predictions. Bollen *et al.* (2011) presents stock market prediction base on Twitter moods. The public mood states are analysed using hypothesis investigated using fuzzy neural network.

A Markov–Fourier grey model that includes a grey model, fourier series and a markov state transition is used by Hsu *et al.* (2009) to predict turning time of a stock index for increased forecasting accuracy. Markov chains is a random process that transition between states that depend their immediate predecessors. Bauerle and Rieder (2004) uses a markov chain based model with stochastic control methods to maximize the expected utility from stock market.

1.1.2   Stock Price Discreteness

Apart from all the aforementioned research, there is a very wide interest on the factors that affect the stock price changes. Harper (2013) states that "there is no clean equation that tells us exactly how a stock price will behave" while presenting likely forces that move the stock price up and down constantly, categorized based on 'fundamental factors' (*fundamental values of the company such as profit*), 'technical factors' (*external factors such as inflation, market trends*) and 'market sentiment' (*psychological factors*). However, over time, many discussions have been made about the influence of 'order size' towards the stock price change with an underlying 'supply-and-demand'.

The changes in stock price occur in two ways; one by decreasing the stock price and the other by increasing the stock price. Therefore, it is important to consider both the increase and decrease of the stock price when analyzing changes of the stock price. This phenomenon seems interesting enough to investigate the stock transactions from a novel angle to provide a different way of understanding how the stock price changes with respect to the order size as a mixture of 'stock price increments' and 'stock price decrements'.

With the stock market regulations, a 'tick size' is maintained to provide a minimum amount by which a stock price can change (*tick size*, 2013). According to Woehrmann (2007) 'tick-size' has been mandated by electronic exchanges as the 'smallest currency unit'. The 'tick-size' governs an indirect discreteness to the changes in stock price where at any given time a stock price can be changed only as a multiple of the established 'tick-size'. This allows the stock prices to cluster among a smaller set of values, instead of taking all the possible real values. Harris (1991) explains that this discreteness and clustering add an advantage towards lowering the cost of negotiation and limiting the information exchange between buyers and sellers. Harris also emphasizes that this clustering is subject to the price level and the volatility of the market.

Due to the discrete nature of stock prices, it is more interesting to look at the changes in stock price as a set of discrete values.

1.2   Introduction to Generalized Linear Models

The equation (1.1) represents a linear regression model where $x_1, \ldots x_k$ are the predictor or regressor variables, y denotes the response or outcome which depends on the predictors, $\beta_1, \ldots \beta_k$ are the unknown parameters to estimate and ε is the random error term.

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k + \varepsilon \qquad (1.1)$$

The classical assumption of the linear regression model is that the response variable (y), takes continuous values following a Normal distribution. However, there are many instances where the response variable (y) may not be continuous. There are instances where the response variable (y) takes binary values such as 'successes' or 'failures', or it could take non-negative integer values such as 'number of defects'. In these cases, the response variable (y) is not a continuous variable and does not follow a normal distribution. Then the above mentioned classical assumption will be violated and a more general regression model will be required.

1.2.1   Generalized Linear Model

The Generalized Linear Model (GLM) was introduced to handle the aforementioned circumstances to allow the response variable (y) to model other types of data and follow distributions other than the normal distribution, as an extension of the linear regression model. Specifically GLM allows the response variable (y) to follow any distribution from the 'exponential family' of distributions, which includes *binomial, normal, poisson, negative binomial, gamma, exponential and inverse normal.* GLM was first introduced by John Nelder and Robert Wedderburn in 1972.

Extending the linear regression model, there are three specifications in using the GLM; they are specification of 'linear predictors', a response variable and use

of a 'link function'.

1. Linear Predictors :

   GLM uses the predictor variables $(x_1, \ldots, x_k)$ as a linear combination of the parameters of the model $(\beta_1, \ldots, \beta_k)$. Let η denote the linear predictors, so that η can be written as $\eta = x_1\beta_1 + x_2\beta_2 + \cdots + x_k\beta_k$.

   This can also be written using the matrix form $\eta = x^T\beta$.

2. Response variable (y):

   Under GLM setting the response variable follows a distribution from the '*Exponential Family*'. Let μ be the mean parameter of the response variable (y). Then $E(y) = \mu$

3. Link function : $g(\cdot)$

   The 'link function', $g(\cdot)$, links the mean parameter, $(E(y) = \mu)$ of the response variable with the linear predictors (η). The link function is a monotonic and one-to-one function with an inverse. Using the monotone and one-to-one properties of the link function the mean can be expressed as a function of linear predictors.

$$
\begin{aligned}
\eta &= g(\mu) \\
\mu &= g^{-1}(\eta) \\
\mu_i &= g^{-1}(x_i^T\beta) \\
\mu_i &= g^{-1}\left(\sum_j x_{ij}\beta_j\right)
\end{aligned}
\qquad (1.2)
$$

Equation (1.2) shows how the mean parameter (μ) is expressed as a function of the linear predictors.

1.2.2 Properties of a Poisson Random Variable

For a random variable with count data or non-negative integers the most natural distribution to use is the 'Poisson distribution'. Poisson distribution is also a member of the 'Exponential Family'. The poisson random variable X, with the mean λ, has following properties.

1. $X \in \{0, 1, 2, 3, \dots\}$

2. The probability function of X : $P(X = x) = \frac{e^{-\lambda}\lambda^x}{(x!)}$

3. The mean and the variance of Y are equal. $E(X) = Var(X) = \lambda$

1.2.3 Poisson Regression Model

The 'exponential family' of the GLM consists of *Natural* and *Exponential Dispersion* sub families. Poisson distribution belongs to the 'Natural' exponential family, whereas the normal distribution belongs to the 'Exponential Dispersion' family. A poisson regression model (PRM) is an example of a GLM where the response variable (y) takes non-negative integers.

The distributions of the 'natural' exponential family can be expressed as the form given in the equation (1.3).

$$f_y(y; \theta) = \exp(y\theta - \kappa(\theta))c(y), \ \theta \in \Omega \tag{1.3}$$

where κ is the cumulant generator, θ is the canonical parameter, and Ω denotes the parameter space which is a subset of the real line.

The probability distribution function of the poisson distribution can be expressed as the form of the 'natural' exponential family given in equation (1.4).

$$f_Y(y; \mu) \qquad = \frac{e^{-\mu}\mu^y}{(y!)}$$

$$f_Y(y; \mu) \quad = \exp\{y\log(\mu) - \mu\} \left(\frac{1}{y!}\right) \tag{1.4}$$

9

When comparing the equation (1.4) with the equation (1.3), the cumulant generator $\kappa(\theta) = \exp(\theta)$ and $\theta = \log(\mu)$. The link function that satisfies $\theta = g(\mu) = \eta = x^T\beta$ is called the 'canonical' or 'natural' link function. Canonical link functions are unique to a particular probability density function (Bonate 2011). Since $\theta = \log(\mu)$ and with the logarithm link function ($\log(\mu) = g(\mu) = \eta$), the link function is 'natural' or 'canonical' for poisson distribution.

Then the logarithm link function is used to link the linear predictors $\eta = x_1\beta_1 + x_2\beta_2 + \cdots + x_k\beta_k$ to the mean parameter of the response variable Y.

$$g(\mu) = \eta$$

$$\log(\mu) = \eta$$

$$\log(\mu) = x_1\beta_1 + x_2\beta_2 + \cdots + x_k\beta_k$$

$$\log(\mu_i) = \sum_j x_{ij}\beta_j$$

The logarithm link function is monotonic and one-to-one with an inverse, therefore the mean parameter of the response variable $(E(Y) = \mu)$ can be expressed as a function of the linear predictors as $\mu_i = e^{\sum_j x_{ij}\beta_j}$ With that the random variable $Y_i \sim Poisson(e^{\sum_j x_{ij}\beta_j})$.

1.2.4   Parameter Estimation

The method of Maximum Likelihood is the theoretical basis for parameter estimation in GLM. This is the most commonly used method to estimate parameters in GLM. According to Madsen & Thyregod (2011), for the GLM with the canonical link function, the likelihood function is convex. Therefore, the maximum likelihood estimate is unique when exists. Then to find the maximum likelihood estimate, one of the most celebrated algorithm called the 'Newton Raphson Algorithm' is used.

1.3 Mixture Models

In analysis of data, often there are heterogeneous sub populations found within the main population. In such cases use of a single distribution will not accurately model the data. It requires modeling the sub populations in order to appropriately model the entire population. The idea of a 'mixture of models' is used to model the populations with heterogeneous sub populations. The history of the 'mixture models' goes back to a classical paper by Karl Pearson in 1894 (McLachlan and Peel 2000). Since then, there have been a wide variety of research performed in many different disciplines including astronomy, biology, genetics, medicine, economics, engineering, marketing and many more.
The probability function of the data with K sub populations is given as

$$f(x;\theta) = \sum_{k=1}^{K} p_k f_k(x;\theta_k)$$

where θ denotes the set of parameters of the mixture model to be estimated and p_k the weight of each probability density function that satisfies $\sum_{k=1}^{K} p_k = 1$ and $0 \leq p_k \leq 1$.

Sometimes the number of sub populations in a population is not a known number. In such cases where the number of sub populations is unknown, the problem leads to an interesting instance of 'estimation of mixture components' which adds extra complexity to the problem. When the number of mixture components is known and finite the problem leads to a 'Finite Mixture Model'.

The weight of each probability density function p_k is also known as the 'mixture probability'. Mixture probability denotes the probability of the observations belonging to each sub population. This being a value between 0 and 1 the p_k could be a constant or an expression of a variable of interest.

1.3.1 Parameter Estimation

Parameter estimation of the mixture models is also performed with the use of Maximum Likelihood (ML) estimation. In estimating parameters for the mixture model there is some additional information required to fulfill in order to estimate the parameters. It is important to know that the data do not identify their sub population. The identification of the sub population pertaining to each observation is missing information in this case. Therefore, ML estimation needs an additional support to fill this missing information. The Expectation and Maximization (EM) algorithm is used to fill the missing information.

DEFINITION 1.1. *The likelihood function for independent and identically distributed data is a product of the densities of the observed value. Thus,*
$L(\theta) = \Pi_{i=1}^{n} f(x_i; \theta)$ *is the likelihood function where θ denotes the unknown parameters and n is the number of observations.*

The likelihood function is a function of the parameters expressed using the observed data. In other words, for the observed data $x_1, \ldots, x_n$ the likelihood function is $L(\theta) = f(x_1)\dot{f}(x_2)\ldots f(x_n)$.

For mixture models the likelihood function is expressed similarly using the distributions of each mixture. For k sub populations and n number of observations, the likelihood function is given by equation (1.5).

$$L(\theta) = \Pi_{i=1}^{n}\Pi_{k=1}^{K} \left[p_k f_k(x_i; \theta_k)\right]^{\Delta_{ik}} \tag{1.5}$$

where Δ_{ik} are unobserved indicators of k sub populations.

However, once the sub populations and their mixture probabilities are included in the likelihood function, the estimation becomes more challenging with the complexity of the likelihood function. It is very common to use the logarithm of the likelihood function, which is known as the 'log-likelihood' to further simplify before maximizing it.

The equation (1.6) below gives the log-likelihood of the likelihood function given in the equation (1.5).

$$l(\theta, p) = \sum_{i=1}^{n} log\,[f(x_i)] = \sum_{i=1}^{n} log\left[\sum_{k=1}^{K} p_k f_k(x_i; \theta_k)\right] \tag{1.6}$$

where $\theta = (\theta_1, \ldots, \theta_k)$ and $p = (p_1, \ldots, p_k)$

The next task is differentiating the 'log likelihood' with respect to each parameter in order to maximize. Then the 'Newton Raphson Algorithm' is used to find the maximum likelihood estimates.

Since the first use of a mixture of two normal distributions by Karl Pearson in 1894 (McLachlan and Peel 2000), there has been a large amount of research performed on mixture models evolving both theory and applications in many different fields. Part of the work on mixture models focuses on identifying the number of components for the mixtures, whereas the remaining works with known number of components.

Brame *et al.* (2006) proposed a mixture model for event count data in criminology to examine the analytical characteristics of the finite mixture models. The model selection criteria Akaike Information Criterion (AIC) and Bayesian Information Criterion(BIC) have been used to decide the appropriate number of mixture components in the mixture model. The authors further confirm that AIC outperforms BIC when the components are very well separated.

According to McLachlan and Peel (2000), AIC selects the models that minimizes the expression AIC$= 2k - 2\ln(L)$, where k is the number of parameters in the mixture model, and L is the maximized value of the likelihood function for the estimated model, and BIC selects the models that minimizes, BIC $= -2 \cdot \ln L + k \ln(n)$ where n is the number of observations

Apart from the applications the theories for mixture models have also been a subject of discussion during last few decades. The optimal rate of convergence

for estimating mixture distributions which consist of finite number of components has also been studied (Chen, 1995). Chen (1995) has proved that when the exact number of the components are known, a consistent rate of $\sqrt{n}$ is achieved and the rate is $n^{-1/4}$ when the number of components are unknown, where n is the sample size.

1.4   EM Algorithm

The Expectation and Maximization (EM) algorithm (Dempster *et al.* 1977) has been a popular tool for simplifying difficult maximum likelihood problems. It is widely used to model applications with missing data or hidden parameters. EM is an iterative algorithm which is used to calculate Maximum Likelihood estimates. Due to various advantages, EM is widely used with mixture models which accommodates a likelihood approach. Yao (2013) states that finding ML estimates in mixture models is an important application of EM.

There are two important steps in EM algorithm; Expectation step (E-step) which calculates the expected probability for the missing probabilities and the Maximization step (M-step) that replaces the missing probabilities with the expected probabilities for maximizing the likelihood function.

1.4.1   EM Algorithm in General

1. Start with initial guesses for the parameters.

2. Expectation Step : calculate the conditional expected value.

3. Maximization Step :

   - fill missing data by the conditional expected values.
   - maximize the simplified likelihood.

4. Iterate expectation and maximization steps until convergence.

EM is most commonly used in mixture models when the parameters are estimated using maximum likelihood method. EM is used to fill the missing information in identifying the mixtures using Δ_{ij} indicators (also known as realizations). When using EM, the likelihood function of the mixture model becomes the 'complete likelihood'.

$$L(\theta) = \Pi_{i=1}^{n} \Pi_{k=1}^{K} f_k(x_i; \theta_k)^{\Delta_{ik}} p_k^{\Delta_{ik}} \tag{1.7}$$

where Δ_{ik} are unobserved realizations of k component indicators. The 'complete log-likelihood' of the equation (1.7) is given by the equation (1.8).

$$l_{complete} = \sum_{i=1}^{n} \sum_{k=1}^{K} \Delta_{ik} \log \left[f_k(x_i; \theta_k) \right] + \sum_{i=1}^{n} \sum_{k=1}^{K} \Delta_{ik} \log[p_k] \tag{1.8}$$

E-step calculates the conditional expected values denoted by (γ_{ij}) to fill the missing information in Δ_{ij}.

$$\gamma_{ij} = E(\Delta_{ij}|p, x) = \frac{f_j(x_i; \theta_j)p_j}{\sum_{k=1}^{K} f_k(x_i; \theta_k)p_k}$$

During the M-step, the Δ_{ij}s in the complete likelihood are replaced with their expected values γ_{ij} and then maximized with respect to each parameter. Then the Newton Raphson algorithm is used to estimate the parameters.

Since the first introduction in 1977, there is a substantial research carried out in using as well as improving the efficiency of the EM algorithm. Xu and Jordan in 1996 have made a comparison of the EM algorithm with several other gradient based algorithms. Several important advantages have been outlined for the EM applied to mixture models. EM is a first order algorithm and under mild conditions it guarantees to converge to the local maximum of the log likelihood. Its naturalness in handling probabilistic constraints of the mixture models also assures the definite convergence compared to algorithms like Newton Raphson. Its low computational overhead has been another reason for its popularity of use.

As much as it is popular due to the advantages, EM is also well known as a slow algorithm in terms of convergence. However, Xu and Jordan (1996) outline that EM performs faster in the mixture setting, especially when the mixture components are well separated. The convergence of EM becomes slower when the mixture components are not well separated. In this situation other gradient based algorithms such as Newton Raphson also do not perform any better due to poor conditions. The authors even mention that under appropriate conditions EM approximates super-linear methods (*a sequence converges superlinearly if it converges and the limit of the rate of convergence is equals to zero*). EM is numerically stable and generally performs well even in ill-conditioned problems. Berlinet and Roland (2012) mentions that EM has a nice convergence and a low cost of memory space. However, a complaint has been made that even though EM iterates quickly to a neighborhood of local model, the final convergence can be very slow and takes a significant amount of time to satisfy the stopping criteria.

As a solution of improvement over the time, Berlinet and Roland (2012) have introduced a newer version of the EM algorithm named as 'Parabolic EM algorithm' (PEM) using the concept of a Beizer Parabola. Their numerical results show that a mixture of two poisson distributions will run 24.5 times faster on the PEM than the regular EM algorithm and the CPU time and failures is 87 : 2138.

A solution to overcome reaching the local maximum is also presented by OHagan *et al.* (2012), especially focusing on optimizing starting values and carefully utilization of maximization steps efficiently.

Selection of appropriate starting values has also been studied as a solution of improving efficiency of the EM algorithm. Wu (1983) states that the convergence of EM algorithm depends on the choice of the starting value. Karlis and Xekalaki (2003) state appropriate initial values speed up the convergence of the algorithm as well as the ability to locate the global maximum. The above authors

have presented a comparison of different algorithms to choose the initial values for the EM algorithm. The results show that the methods such as selection of the moments of the parameters and symmetric method are outperformed by the method introduced by Finch *et. al* in 1989.

1.5 Applications of Mixture Models

The attractions of the mixture models especially increased during recent decades with a wide range of applications from biology, medicine, genetics, business and many more.

A mixture of poisson distributions has been used in segmentation and value estimation in PET (Positron Emission Tomography) data with the use of EM algorithm (Su *et al.* 2011). "PET is a test that uses a special type of camera and a radioactive chemical to look at organs in the body" (*PET*, 2011). According to the authors this is the first time use of the EM based mixture of poisson distributions in the medical field. A comparison of a mixture of Gaussian distributions with the mixture of the poisson distributions has resulted that the poisson mixture provided more robust and accurate results. It is also postulated by the authors, due to the fact that the poisson distribution has a single parameter where as normal distribution has two might have made the convergence faster.

Finite mixtures of poisson and negative binomial distributions have been used to model vehicle crash data in 2009 by Park and Lord (Park and Lord, 2009). Instead of using the maximum likelihood approach with EM algorithm the authors have used a 'Bayesian sampling approach'.

In 2008 a mixture of poisson distributions were used in an application of identifying changes in RNA polymerase II binding quantity using high-throughput sequencing technology (Feng *et al.* 2008). It has been assumed that the number

of 'poly II targeted sequences' contained in each genomic region follows a poisson distribution. The proposed mixture of poisson distributions were used to distinguish 'Pol II binding changes in transcribed region' with the use of EM algorithm. Two types of breast cancer cells were modeled using the mixtures.

Mixture models is also used in clustering under data mining, where a dataset needs to be grouped into unknown set of subgroups using patterns and relationships among data. This instance is similar to the instance of a mixture model where the number of sub populations are unknown. In 2004 a similar model was used for an application designed for retail category managers to improve customized merchandising strategies (Brijs *et al.* 2004). The application was used to cluster supermarket shoppers based on their purchase patterns. A multivariate poisson mixture model has been used among the other techniques.

A mixture of normal distributions were used to classify dolphins found in California Bight and the Gulf of California in 2007 (Rocha *et al.*, 2007). The researches have monitored vocal signals of short-beaked common, pacific white-sided and bottle nose dolphins for a period of four years. Then these vocal signals were processed and used with a mixture of normal distributions with an unknown number of components. The testing has resulted in a 256 component model over mixture models of 64, 128 and 512 components, since it is more accurate in identifying the different vocal signals to help classifying dolphins.

Text mining is extracting knowledge and identifying patterns among text documents. Mixture models have also been useful in classifying text patterns and identifying themes across different articles. Zhai *et al.*, (2004) have performed research on a 'Cross Collection Mixture Model for Comparative Text Mining' to summarize, compare and contrast the common themes across multiple news articles. They have used datasets with news articles and laptop reviews to discover common themes across the collection of articles and summarize their similarities

and differences along each common theme.

Mixture models have also been used for applications in identifying technology characteristics. One instance is to characterize end-to-end Internet delay (Hernandez and Phillips, 2006). A finite mixture model is used to characterize and summarize the delay experienced between edges of the Internet, using Weibull distributions. The Weibull distribution is a continuous distribution which is used to model lifetimes of objects. Real GPS synchronized measurements have been used to validate the model.

Gerdtham and Trivedi (2001) presents a finite mixture model to analyze the utilization of Swedish health care. A 2-component negative binomial distribution has been used to distinguish the utilization of health care facilities by patients of different social and income groups.

Applications of finite mixture models are also found in image processing. According to Blekas *et al.* (2005) one of the most successful applications of Gaussian mixture models is found in image segmentation. The authors have implemented an improved version of the M-step of the EM algorithm and have used models with 3 and 5 component Gaussian mixtures to evaluate the improvements. A similar application of a finite mixture model is found in face-color modeling and segmentation (Greenspan *et al.*, 2001) which uses a mixture of Gaussian distributions. The mixture is used to model the color space to provide a robust representation while permitting a large color variations, highlights and shadows.

Another interesting application of mixture models is found in evaluating "Which Micro-finance Institutions Are Becoming More Cost Effective with Time?" according to a research published by Caudill *et al.* in 2009. A mixture of two normal distributions is applied to the data on micro finance institutions from Eastern Europe and Central Asia. The authors have found that larger micro-finance institutions that offer deposits and the micro-finance institutions that receive lower

subsidies operate more cost effectively over time.

1.6   Data Set

Federal Express Corporation which is widely known as FedEx (FDX) was incorporated in 1971 as an overnight delivery service. FDX sold their first shares on the New York Stock Exchange in 1978 (*History of FedEx Operating Companies*, 2013). Since then FDX holds many success stories including being one of the seven super star companies to have several stock price splits, according to Moroney (2007). FDX have had five stock price split throughout their history. A stock price split shows an added advantage to a company denoting that the company's financial stability and trend to attract more individual investors.

The data set under investigation consists of 'tick-by-tick' daily stock transactions of five years from FDX. One interesting fact about this dataset is, it consists of the company's fourth stock price split. Each transaction consists of the logarithm stock price, the order size of the transaction and the time of the transaction within the day. This dataset is used to understand the changes which occur in the stock price.

1.7   Chapter Outline

Data preprocessing is regarded as an important task, when working with data. It even helps successful handing of row data when it is noisy, incomplete and inconsistent. This leads to minimize errors and provides a better understanding of the data. Therefore, the five years of stock transactions data is first preprocessed and then the discrete stock price changes are identified. This is described in Chapter 2.

Using the discrete stock price changes, a mixture model is proposed in

chapter 3 under conditions where first as a constant mixture probability followed by a mixture probability as a function of the order size.

Chapter 4 presents different attempts and results of efficiency improvements to the model including the implementation of the 'Parabolic EM' algorithm. A 'Clustered Signed Model' is presented to utilize less amount of data by summarizing the large amount of data that the proposed mixture model uses and is presented in chapter 5.

A test is proposed to verify the suitability of the mixture probability as a function of the order size, in chapter 6. Chapter 7 presents approximated confidence intervals for the parameter estimates of the model using a simulation.

Chapter 8 presents the discussion, conclusions and the future of the current research.

CHAPTER 2
DATA ANALYSIS

2.1   Understanding Data

First of all it is important to analyze the data set thoroughly to understand and prepare them for effective use. In this chapter the data pre-processing and preparation for the implementation is discussed in detail.

The Federal Express (FDX) data set consists of the daily transactions of stocks of the company. The data set contains tick-by-tick stock transactions for 1263 days during five years. That includes 238575 tick-by-tick transactions. Each transaction consists of log values of the stock price, volume of the transaction (*aka* trade size) and the time of the transaction in seconds within trading the day.

Figure 2.1 shows the plot of the log stock price of the entire data set. A sample of the original data is given in Appendix A.1. As the first task, it was required to convert the log stock price into the original stock price. The stock price was obtained by exponentiating the log stock price. During this calculation precision of the stock price is lost and thus it was needed to approximate the nearest tick-size. Therefore, an additional step was required to restore the accuracy of the actual stock price, approximating the values to the closest tick-size. Figure 2.2 shows the plot of the actual stock price throughout the five years.

With the tick-size being maintained as the smallest value the stock price change the stock price has become a discrete value. Stock price is at its minimum of
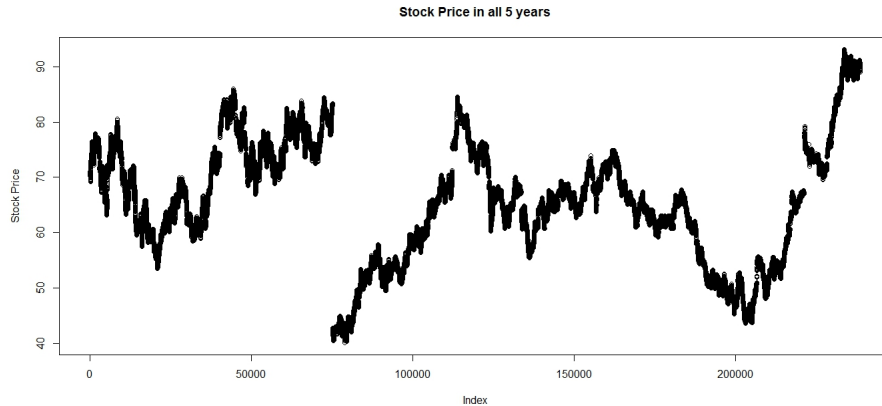
FIGURE 2.1 – Stock price during all 5 years

$40.13 during the 3rd year. This was due to the stock price split on 5th November 1996.

Once the stock price reaches a higher level, companies decide to split the price of their stocks (*What is Stock Split?*, 2009). The stock price split does not change the value of the company. It makes the stocks more affordable to small investors. Stock markets consider a 'price split' as another way of attracting more investors resulting an increased demand of the stock.

According to analysis of the data, the number of transactions have increased after the stock price split as expected by the stock market. The maximum stock price of $93.13 has been observed on 7th December in 1997, about a year after the stock price split. The summary statistics of the stock price during five years are shown in table 2.1. However, the data shows that during 62.8% of the transactions during the five years, the stock price has not changed due to the trade.

Stocks are purchased or sold as multiples of hundreds with the smallest unit of the order size being 100. This makes the order size a discrete variable. The purchases are denoted by positive values in the order size whereas sales are denoted by negative values in the order size. For example, -200 order size means that 200 stocks were sold and 300 order size means that 300 stocks were purchased.

Minimum	1st Quartile	Median	Mean	3rd Quartile	Maximum
40.13	59.00	65.75	65.54	73.31	93.13

TABLE 2.1

Summary statistics of the stock price

Minimum	1st Quartile	Median	Mean	3rd Quartile	Maximum
-500000.0	-500.0	100.0	82.5	600.0	628800.0

TABLE 2.2

Summary statistics of the order size

Different amounts of stocks were sold and purchased during these five years. There were purchases as large as 628800 of stocks and sales as large as 500000 stocks (figure 2.2). However, 92.89741% of transactions included order size 5000 or less and 26.86115% transactions were made with either 100 or 200 stocks.

Among other influential factors, the order size also plays an important role in changing the stock price. Purchases of more stocks make the stock price increase more while sales of more stocks make the stock price decrease more. This phenomenon is known as the 'Market Impact'. According to Moro *et al.* (2009) the 'Market Impact' is the expected price change based on the trades and the trade sizes. Therefore, in analyzing the price change, the order size also needs to be analyzed.

The stock market opens at 9.30am and closes at 4.00pm for the day. During the day a large number of transactions occur. For each transaction the time is recorded in seconds during the day. The stock transactions during the five years were analyzed based on the time of the day. A significantly large number of the
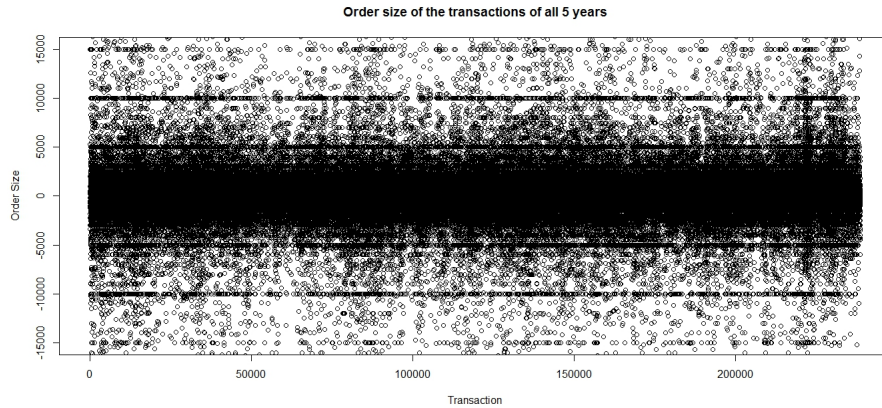
FIGURE 2.2 – Order size
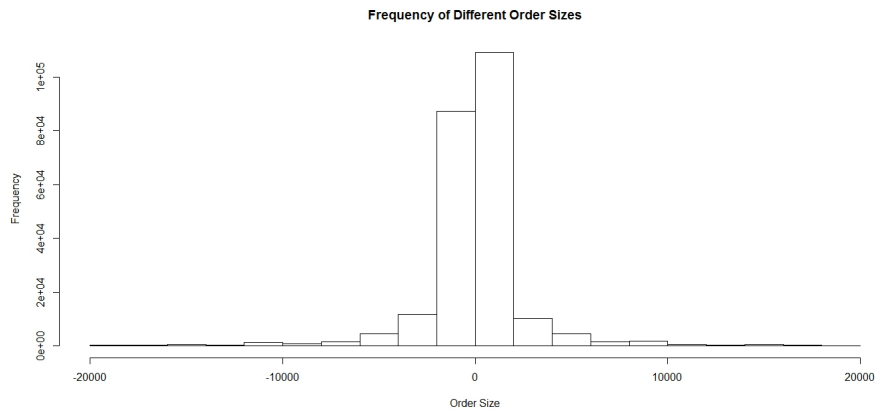
FIGURE 2.3 – Frequency of Different Order Sizes

transactions occur between 10.00am to 10.30am in the morning and 3.30pm to 4.00pm in the afternoon. This has been consistent throughout all five years. Out of those more transactions were found between 10.00am to 10.30am during first four years and more found in 3.30pm to 4.00pm during the last year. This could be explained by the stock price split at the end of the 3rd year and the tick size change to 1/16 during the middle of the fourth year. Figures 2.5 and 2.6 show the frequency of the transactions during the day during all five years.

Figures 2.7 to 2.11 show the plots of stock price during each year. Figure 2.7 shows an overall decline in the price towards the end of the year and continues

25
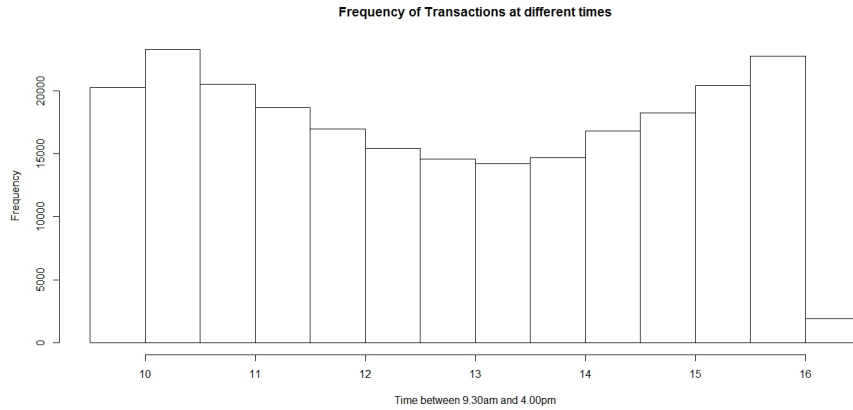
FIGURE 2.4 – Frequency of Transactions during the day
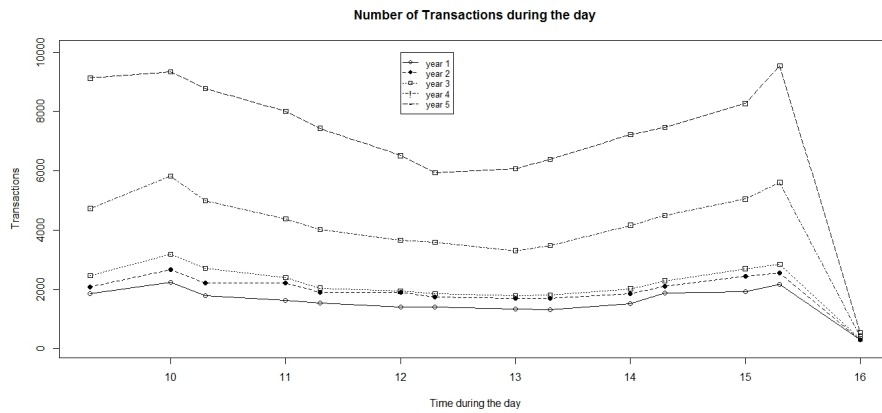
FIGURE 2.5 – No. of transactions during the day of five years

to the beginning of the year two as shown in figure 2.8. However, the stock price starts to increase at the middle of the year two and shows an overall increase by the end of year. The stock price kept consistently high during the third year. As a result the company employed the stock price split at the end of the year 3.

The number of transactions also have increased during each year. Year 2 had 22% more transactions than year 1, year 3 had 11% more than year 2, year 4 had 89% more than year 3 and the fifth year had 74% more than the year 4. The stock price split at the end of year 3 explains the largest increase of 89% in transactions from year 3 to year 4.

26
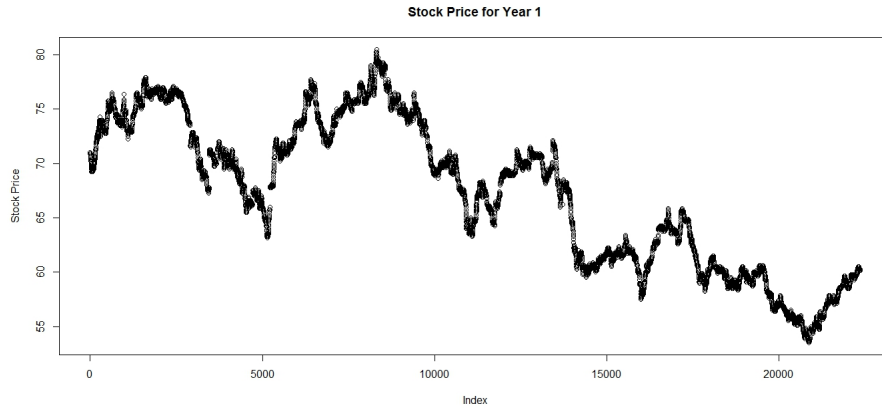
FIGURE 2.6 – Tick-by-tick stock price during year 1
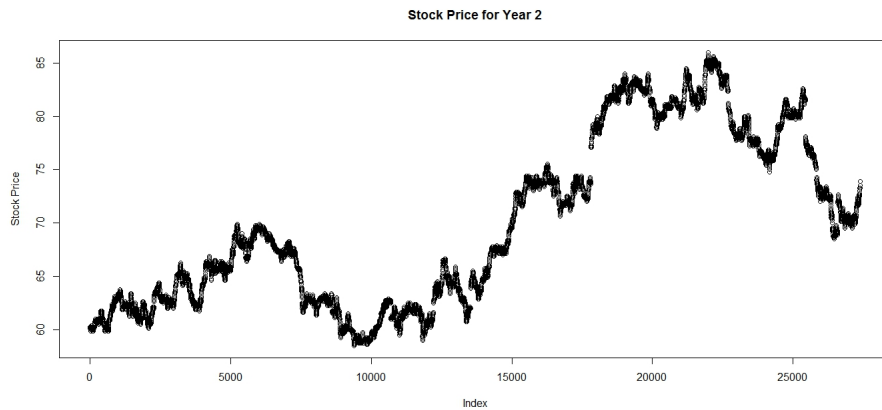
FIGURE 2.7 – Tick-by-tick stock price during year 2

Apart from the stock price split, there is another important change which took place during the five years. That is the change of the tick-size from 1/8 to 1/16. The tick-size at the beginning of the five year year period was 1/8. After 3 1/2 years, on 23rd June 1997 tick-size was changed to 1/16. The stock price was changed by 1/16 during the last 1 1/2 years of the data set. This change in the tick-size is also an interesting factor to consider when analyzing data. Due to the decrease of the tick-size there were 40% more transactions made during last 1 1/2 years than the first 3 1/2 years.

With all the above findings it seems important to investigate the data as

27
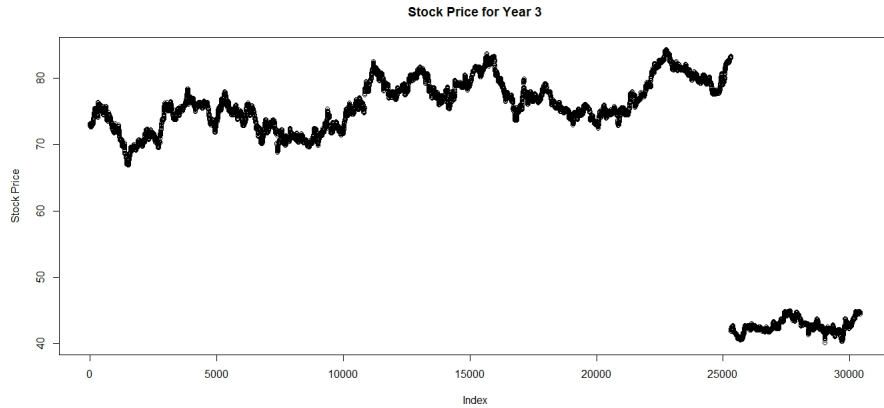
FIGURE 2.8 – Tick-by-tick stock price during year 3
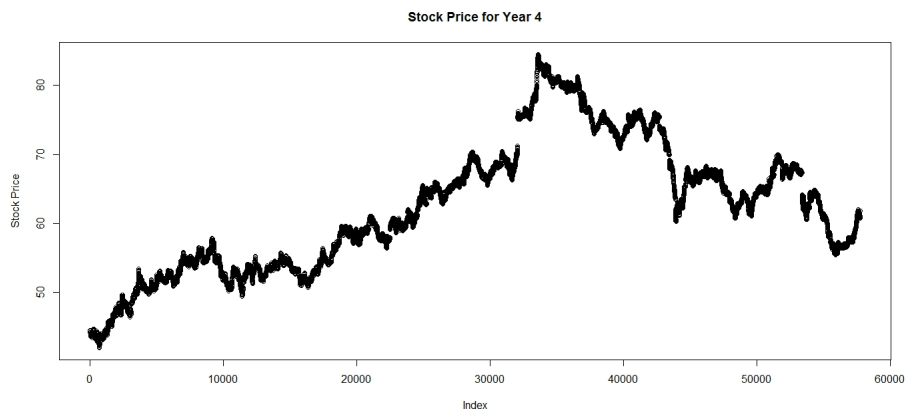
FIGURE 2.9 – Tick-by-tick stock price during year 4

sub sets based on each year as well as based on different tick-sizes. Thus all the investigations throughout were performed for the sub data sets considering the year and the corresponding tick-size.

2.2   Data Preparation for Mixture Model

The focus on this study is to understand the changes to the stock prices. It requires calculating the price changes of the stock prices of each transaction. Based on the tick-size it is possible to identify the changes in the stock price, as a multiple of the tick-size.

28
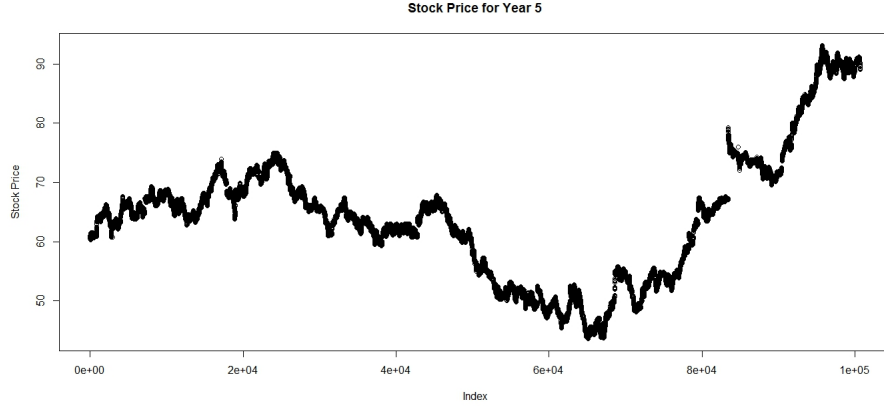
FIGURE 2.10 – Tick-by-tick stock price during year 5

Year	No. of transactions	% Increment of Transactions
year 1	22358	-
year 2	27390	22.5%
year 3	30413	11.0%
year 4	57733	89.8%
year 5	100681	74.4%

TABLE 2.3

Increase of transactions over the years

Let $S(t,x)$ denotes the stock price per share at time t that a trader receives for an order of size x. Given observed settlement prices $S(t_i, x_i)$ for $i = 0, 1, \ldots, n$ at times $t_0 < t_1 < \ldots < t_n$, let

$$Y_i = \frac{S(t_i, x_i) - S(t_{i-1}, x_{i-1})}{\tau} \tag{2.1}$$

be the number of ticks that the stock price moves when the ith transaction is made based on the tick-size τ. Then $Y_i \in \{\ldots, -2, -1, 0, 1, 2, \ldots\}$.

The number of ticks denoted by Y_i takes integer values. The negative sign in

29
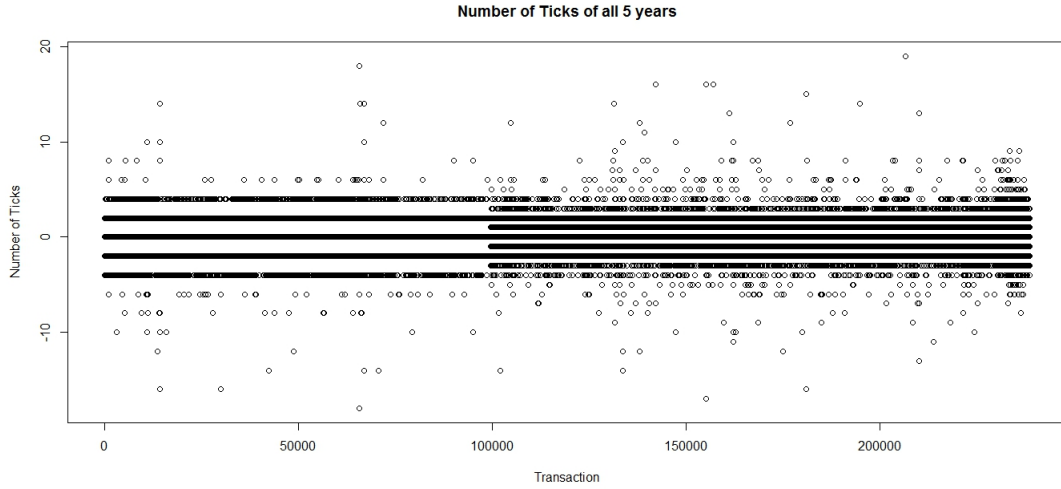
FIGURE 2.11 – Number of ticks of all five years

integers denotes the instances where the price goes down between two consecutive transactions and the positive sign in integers are obtained for those where the price increased between two consecutive transactions. The value zero was obtained when the consecutive transactions did not have any price change. The magnitude shows change in price as a multiple of the tick-size that was used during the time period.

Figure 2.12 shows the plot of number of ticks (Y_i) of all five years. The figure clearly shows the clouds of discrete values as horizontal lines and the separation of tick-size from 1/8 to 1/16 at the middle of the plot. Figures 2.13 and 2.14 show a closer view of the number of ticks separated by tick-size. A sample of processed data is given in Appendix A.2.
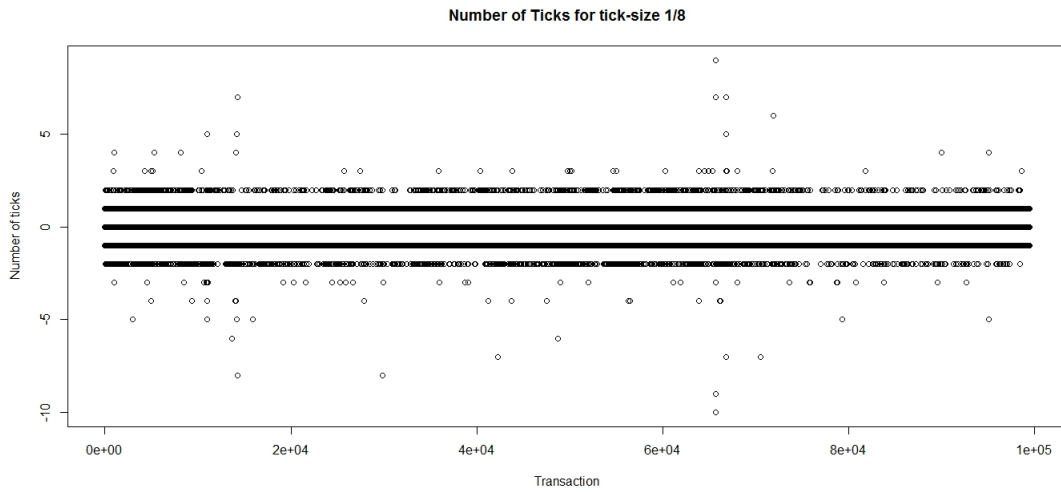
FIGURE 2.12 – Number of ticks for tick size 1/8
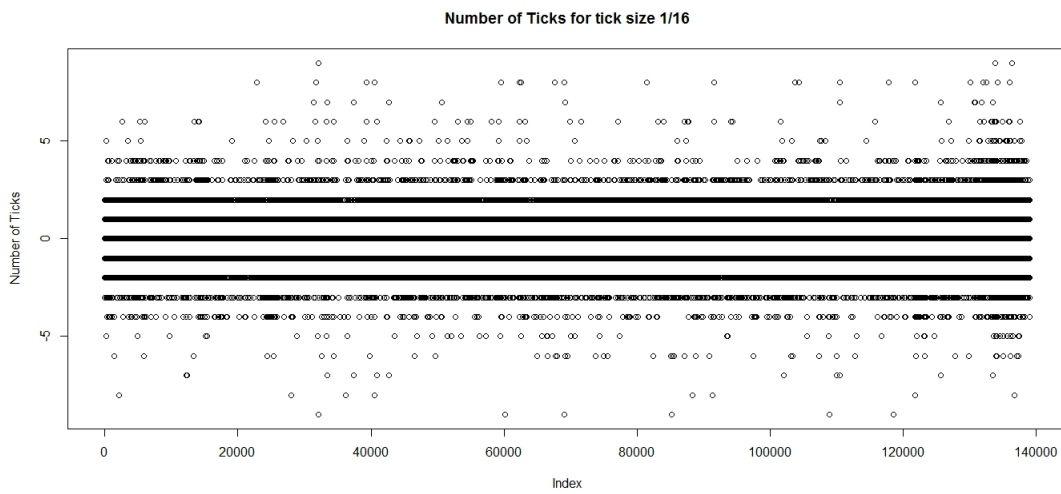
FIGURE 2.13 – Number of ticks for tick size 1/16

31

CHAPTER 3
THE MODEL

3.1   Model Formulation

Chapter 2 presented the preparation of FDX stock transactions data which is ready to use for the model formulation. The rest of the chapter discusses on formulating model.

The main interest is to investigate 'How the stock price changes with respect to the purchases and sales of stocks'. This determines the response variable to be the 'change in the price' with respect to number of predictor variables. The stock price changes occur discretely. Therefore, the response variable is not continuous and violates the normality assumption of the simple regression model. As it is discussed in chapter 1, this is when it is required to use the 'Generalized Linear Models' (GLM). GLM requires three important things to specify. This includes the response variable, the linear predictors and the link function.

3.1.1   Response Variable is a Mixture

The response variable is the number of ticks denoted by Y. Y consists of integers. Thus $Y \in \{\ldots, -2, -1, 0, 1, 2, \ldots\}$. Here, the negative number of ticks denote the stock price decrements and the positive number of ticks denote the stock price increments. The value 0 was obtained from the transactions where the price did not change between them.

The number of ticks consists of a mixture of stock price increments and stock price decrements. There are two sub populations within the number of ticks, where each sub populations is represented by its corresponding distribution. This leads to the possibility of proposing a mixture model.

Define a random variable Y_1 that contains the non-negative integers of y_i. Let $Y_1 \in \{0, 1, 2, 3, \dots\}$. With Y_1 containing non-negative integers, the first natural guess is that considering Y_1 as a poisson random variable. Then let Y_1 follows a poisson distribution with mean parameter λ^+. Then define another random variable Y_0 containing the non-positive integers of y_i. Let $Y_0 \in \{\dots, -3, -2, -1, 0\}$. Next negate the non-positive random variable Y_0. That results the random variable denoted by $-Y_0$ which takes only non-negative integers. Thus $-Y_0 \in \{0, 1, 2, 3, \dots\}$. Similarly, it can be stated that $-Y_0$ follows a poisson distribution with mean parameter λ^-.

The number of ticks formulated based on both stock price increments and stock price decrements are modeled using the two random variables Y_1 and Y_0. Thus the mixture of number of ticks calculated based on stock price increments and decrements can now be modeled using a mixture of two poisson random variables.

$$Y_1 \sim Poisson(\lambda^+)$$

$$-Y_0 \sim Poisson(\lambda^-)$$

It can be seen that the random variable Y consists of the values from the two random variables Y_1 and Y_0. In order to specify the relationship of Y_1 and Y_0 within Y, it requires a need of an indicator variable. Let Δ be the indicator variable which takes binary values. Then Δ is a 'Bernoulli' random variable and will denote the membership of each observation in each sub population. Let p be the probability for $\Delta = 1$. Then the mixture of two poisson distributions can be specified as below. The proposed model assumes that y_i is an observed value of

the random variable Y which is a mixture of two poisson random variables Y_0 and Y_1.

$$Y_1 \sim Poisson(\lambda^+)$$

$$-Y_0 \sim Poisson(\lambda^-)$$

$$Y = (1 - \Delta) \cdot Y_0 + \Delta \cdot Y_1$$

$$\text{where } \Delta \in \{0, 1\} \text{ with } Pr\{\Delta = 1\} = p$$

3.1.2   Linear Predictors - Order Size

The next task is to identify the predictors. The stock transactions data from FDX consists of the trade size and the time of the transaction within the day. As order size plays an important role in stock price movements, it was decided to start with 'order size' as the predictor variable. Then the linear combination of the order size can be written using the equation (3.1). Here x_i denotes the order size of the ith transaction.

$$\eta_i = \beta_{0i} + \beta_{1i} x_i \tag{3.1}$$

The response variable consists of a mixture of two random variables. Therefore, the linear predictors are expressed for each variable.

$$\text{for } Y_1 : \eta_{1i} = \beta_{0i}^+ + \beta_{1i}^+ x_i$$

$$\text{for } Y_0 : \eta_{0i} = \beta_{0i}^- + \beta_{1i}^- x_i$$

3.1.3   Link Function

The logarithm link function is used to link the mean parameter of the response variable Y with the linear predictors η. The response variable is a mixture of two random variables and each has a mean parameter. Therefore, the link

function is applied to the mean parameters of Y_0 and Y_1.

$$\text{for } Y_1 : \log(\lambda_i^+) = \beta_0^+ + \beta_1^+ x_i \Rightarrow \lambda_i^+ = e^{\beta_0^+ + \beta_1^+ x_i}$$

$$\text{for } Y_0 : \log(\lambda_i^-) = \beta_0^- + \beta_1^- x_i \Rightarrow \lambda_i^- = e^{\beta_0^- + \beta_1^- x_i}$$

The proposed model can be completely specified as below.

$$\Delta_i \sim \text{Bernoulli}(p), \ i = 1, \ldots, n$$

$$y_i | \Delta_i \sim \text{Poisson}(\lambda_i^+ = e^{\beta_0^+ + \beta_1^+ x_i})$$

$$-y_i | \Delta_i \sim \text{Poisson}(\lambda_i^- = e^{\beta_0^- + \beta_1^- x_i})$$

As the initial case of investigation the probability of $\Delta = 1$ is considered as a constant. That is $P(\Delta = 1) = p$ where p is a constant. An extension to this case where p is a function of the order size will be discussed in the next section.

3.1.4 Maximum Likelihood Method

Parameters are estimated using the maximum likelihood estimation. The likelihood function for the mixture of two sub populations can be expressed as given in equation (3.2) with $\theta = (p, \beta_0^-, \beta_1^-, \beta_0^+, \beta_1^+)$.

$$L(\theta) = \Pi_{i=1}^n [(1-p)f_1(y_i)]^{1-\Delta_i}[pf_2(y_i)]^{\Delta_i} \tag{3.2}$$

With the underlying poisson distribution the probability function become the probability mass function of the discrete poisson distribution. The function 'P' denotes the probability mass function of poisson distribution in the equation (3.3).

$$L(p, \beta_0^-, \beta_1^-, \beta_0^+, \beta_1^+) = \Pi_{i=1}^n [(1-p)P_1(y_i)]^{1-\Delta_i}[pP_2(y_i)]^{\Delta_i} \tag{3.3}$$

The logarithm of the likelihood function, $l(\theta)$, is obtained to make the calculations easier.

$$
\begin{aligned}
l(p, \beta_0^-, \beta_1^-, \beta_0^+, \beta_1^+) &= \sum_{i=1}^{n} \ln P(y_i) \\
&= \sum_{i=1}^{n} \ln\{P(y_i|\Delta_i = 0)P(\Delta_i = 0) + P(y_i|\Delta_i = 1)P(\Delta_i = 1)\} \\
&= \sum_{i=1}^{n} \ln\{(1-p)\frac{(\lambda_i^-)^{-y_i}e^{-\lambda_i^-}}{(-y_i)!}I_{y_i \leq 0} + p\frac{(\lambda_i^+)^{y_i}e^{-\lambda_i^+}}{(y_i)!}I_{y_i \geq 0}\} \quad (3.4)
\end{aligned}
$$

It is required to identify the indicator variables denoted by '$I_{y_i \leq 0}$' and '$I_{y_i \geq 0}$' in order to maximize the log-likelihood given in equation (3.4). It is a missing information in the model. Therefore, the EM algorithm is used to fill the missing information.

The membership of observations in the two sub populations are identified during the 'E-step'. It is important to identify the value zero belongs to both sub populations and thus the probability of its membership is also required. The 'E-step' calculated the γ_i using the 'conditional probability theorem' and the 'total probability theorem'.

"E" - Step

$$
\gamma_i = P(\Delta_i = 1|y_i) = \frac{P(y_i|\Delta_i = 1)P(\Delta_i = 1)}{P(y_i|\Delta_i = 1)P(\Delta_i = 1) + P(y_i|\Delta_i = 0)P(\Delta_i = 0)} \quad (3.5)
$$

If $y_i < 0$, then $P(y_i|\Delta_i = 1) = 0 \qquad \Rightarrow \gamma_i = 0$

If $y_i > 0$, then $P(y_i|\Delta_i = 1) = 1 \qquad \Rightarrow \gamma_i = 1$

If $y_i = 0$, then

$$
\begin{aligned}
\gamma_i &= \frac{\hat{p}e^{-\hat{\lambda}_i^+}}{\hat{p}e^{-\hat{\lambda}_i^+} + (1-\hat{p})e^{-\hat{\lambda}_i^-}} \\
&= \frac{\hat{p}e^{-e^{\hat{\beta}_0^+ + \hat{\beta}_1^+ x_i}}}{\hat{p}e^{-e^{\hat{\beta}_0^+ + \hat{\beta}_1^+ x_i}} + (1-\hat{p})e^{-e^{\hat{\beta}_0^- + \hat{\beta}_1^- x_i}}}
\end{aligned}
$$

The 'M-step' replaces the Δ_i with γ_i in the log-likelihood which is then called the 'complete log-likelihood' (equation 3.5).

"M" - Step

$$l_0(p, \beta_0^-, \beta_1^-, \beta_0^+, \beta_1^+) = \sum_{i=1}^n (1 - \Delta_i) \ln P(y_i; \beta_0^-, \beta_1^-) + \sum_{i=1}^n \Delta_i \ln P(y_i; \beta_0^+, \beta_1^+)$$

$$+ \sum_{i=1}^n (1 - \Delta_i) \ln(1 - p) + \sum_{i=1}^n \Delta_i \ln p \ (3.6)$$

Then the complete log-likelihood given in equation (3.6) is differentiated with respect to each parameter to obtain the maximum likelihood estimates.

Finding $\hat{p}$:

$$\begin{aligned}
\frac{\partial l_0}{\partial p} &= \sum_{i=1}^n (1 - \Delta_i) \frac{-1}{1 - p} + \sum_{i=1}^n \Delta_i \frac{1}{p} \\
&= \frac{1}{p(1 - p)} \{ -p \sum_{i=1}^n (1 - \Delta_i) + (1 - p) \sum_{i=1}^n \Delta_i \} \\
\Rightarrow \hat{p} &= \frac{\sum_{i=1}^n \Delta_i}{n}
\end{aligned} \tag{3.7}$$

Finding $\hat{\beta}_0^-$:

$$\begin{aligned}
\frac{\partial l_0}{\partial \beta_0^-} &= \frac{\partial}{\partial \beta_0^-} \sum_{i=1}^n (1 - \Delta_i) \ln P(y_i; \beta_0^-, \beta_1^-) \\
&= \sum_{i=1}^n (1 - \Delta_i) \frac{1}{P(y_i; \beta_0^-, \beta_1^-)} \frac{\partial}{\partial \beta_0^-} P(y_i; \beta_0^-, \beta_1^-)
\end{aligned}$$

Since $P(y_i; \beta_0^-, \beta_1^-) = \frac{\exp\{-e^{(\beta_0^- + \beta_1^- x_i)}\}(e^{\beta_0^- + \beta_1^- x_i})^{y_i}}{(y_i)!}$ it results;

$$\begin{aligned}
\frac{\partial P(y_i; \beta_0^-, \beta_1^-)}{\partial \beta_0^-} &= \frac{1}{(y_i)!} (e^{\beta_0^- + \beta_1^- x_i})^{y_i - 1} \exp\{-e^{(\beta_0^- + \beta_1^- x_i)}\} \{y_i - e^{\beta_0^- + \beta_1^- x_i}\} \\
\frac{\partial l_0}{\partial \beta_0^-} &= \sum_{i=1}^n e^{-\lambda_i^-} (1 - \Delta_i) \{y_i - e^{\beta_0^- + \beta_1^- x_i}\} = 0 \\
\Rightarrow \sum_{i=1}^n (1 - \Delta_i) \{y_i - e^{\beta_0^- + \beta_1^- x_i}\} &= 0
\end{aligned} \tag{3.8}$$

Finding $\hat{\beta}_1^-$:

$$
\begin{aligned}
\frac{\partial l_0}{\partial \beta_1^-} &= \frac{\partial}{\partial \beta_1^-} \sum_{i=1}^{n} (1 - \Delta_i) \ln P(y_i; \beta_0^-, \beta_1^-) \\
\frac{\partial P(y_i; \beta_1^-, \beta_1^-)}{\partial \beta_1^-} &= \frac{1}{(y_i)!} (e^{\beta_0^- + \beta_1^- x_i})^{y_i - 1} \exp\{-e^{(\beta_0^- + \beta_1^- x_i)}\} \{y_i - e^{\beta_0^- + \beta_1^- x_i}\} x_i \\
\frac{\partial l_0}{\partial \beta_1^-} &= \sum_{i=1}^{n} e^{-\lambda_i^-} (1 - \Delta_i) \{y_i - e^{\beta_0^- + \beta_1^- x_i}\} x_i = 0 \\
&\Rightarrow \sum_{i=1}^{n} (1 - \Delta_i) \{y_i - e^{\beta_0^- + \beta_1^- x_i}\} x_i = 0 \qquad (3.9)
\end{aligned}
$$

Finding $\hat{\beta}_0^+$:

$$
\begin{aligned}
\frac{\partial l_0}{\partial \beta_0^+} &= \frac{\partial}{\partial \beta_0^+} \sum_{i=1}^{n} \Delta_i \ln P(y_i; \beta_0^+, \beta_1^+) \\
&= \sum_{i=1}^{n} \Delta_i \frac{1}{P(y_i; \beta_0^+, \beta_1^+)} \frac{\partial}{\partial \beta_0^+} P(y_i; \beta_0^+, \beta_1^+) \\
\text{Since } P(y_i; \beta_0^+, \beta_1^+) &= \frac{\exp\{-e^{(\beta_0^+ + \beta_1^+ x_i)}\}(e^{\beta_0^+ + \beta_1^+ x_i})^{y_i}}{(y_i)!} \text{ it results;} \\
\frac{\partial P(y_i; \beta_0^+, \beta_1^+)}{\partial \beta_0^+} &= \frac{1}{(y_i)!} (e^{\beta_0^+ + \beta_1^+ x_i})^{y_i - 1} \exp\{-e^{(\beta_0^+ + \beta_1^+ x_i)}\} \{y_i - e^{\beta_0^+ + \beta_1^+ x_i}\} \\
\frac{\partial l_0}{\partial \beta_0^+} &= \sum_{i=1}^{n} e^{-\lambda_i^+} \Delta_i \{y_i - e^{\beta_0^+ + \beta_1^+ x_i}\} = 0 \\
&\Rightarrow \sum_{i=1}^{n} \Delta_i \{y_i - e^{\beta_0^+ + \beta_1^+ x_i}\} = 0 \qquad (3.10)
\end{aligned}
$$

Finding $\hat{\beta}_1^+$:

$$
\begin{aligned}
\frac{\partial l_0}{\partial \beta_1^+} &= \frac{\partial}{\partial \beta_1^+} \sum_{i=1}^{n} \Delta_i \ln P(y_i; \beta_0^+, \beta_1^+) \\
\frac{\partial P(y_i; \beta_1^+, \beta_1^+)}{\partial \beta_1^+} &= \frac{1}{(y_i)!} (e^{\beta_0^+ + \beta_1^+ x_i})^{y_i - 1} \exp\{-e^{\beta_0^+ + \beta_1^+ x_i}\} \{y_i - e^{\beta_0^+ + \beta_1^+ x_i}\} x_i \\
\frac{\partial l_0}{\partial \beta_1^+} &= \sum_{i=1}^{n} e^{-\lambda_i^+} \Delta_i \{y_i - e^{\beta_0^+ + \beta_1^+ x_i}\} x_i = 0 \\
&\Rightarrow \sum_{i=1}^{n} \Delta_i \{y_i - e^{\beta_0^+ + \beta_1^+ x_i}\} x_i = 0 \qquad (3.11)
\end{aligned}
$$

After simplifications the score equations are :

$$\hat{p} = \frac{\sum_{i=1}^{n} \Delta_i}{n} \qquad (3.7)$$

$$\sum_{i=1}^{n} (1 - \Delta_i)\{y_i - e^{\beta_0^- + \beta_1^- x_i}\} = 0 \qquad (3.8)$$

$$\sum_{i=1}^{n} (1 - \Delta_i)\{y_i - e^{\beta_0^- + \beta_1^- x_i}\}x_i = 0 \qquad (3.9)$$

$$\sum_{i=1}^{n} \Delta_i\{y_i - e^{\beta_0^+ + \beta_1^+ x_i}\} = 0 \qquad (3.10)$$

$$\sum_{i=1}^{n} \Delta_i\{y_i - e^{\beta_0^+ + \beta_1^+ x_i}\}x_i = 0 \qquad (3.11)$$

If Δ_i's are known, the mean of the Δ's give the estimate of the p, as given in (3.7). The estimates of β_0^- and β_1^- can be obtained from (3.8) and (3.9) using a weighted Poisson regression, and the estimates of β_0^+ and β_1^+ can be obtained from (3.10) and (3.11) using another weighted Poisson regression.

Only the estimator for mixture probability ($\hat{p}$), given in equation (3.7), has a closed form. Other equations of $\hat{\beta}_0^-, \hat{\beta}_1^-, \hat{\beta}_0^+, \hat{\beta}_1^+$, given by equations (3.8) - (3.11) do not have closed forms. Therefore the Newton Raphson method is used to find $\hat{\beta}_0^-, \hat{\beta}_1^-, \hat{\beta}_0^+, \hat{\beta}_1^+$.

It can be seen that there is a pair of equations to be solved for each mixture. The pair (3.8) and (3.9) belong to the non-positive counts based on the stock price decrements and the pair (3.10) and (3.11) belong to the non-negative counts based on the stock price increments. It can also be observed that both pairs have the similar parametric formulation.

The EM algorithm can be implemented for this model using the following steps.

1. Take initial guesses $\hat{p}$, $\hat{\beta}_0^-$, $\hat{\beta}_1^-$, $\hat{\beta}_0^+$, and $\hat{\beta}_1^+$ for the parameters.

2. *E–step*: Compute the expected values of each Δ_i using γ_i given in (3.5). If $y_i < 0$ then $\gamma_i = 0$, if $y_i > 0$ then $\gamma_i = 1$ and if $y_i = 0$ then,

$$\gamma_i = \frac{\hat{p}e^{-e^{\hat{\beta}_0^+ + \hat{\beta}_1^+ x_i}}}{\hat{p}e^{-e^{\hat{\beta}_0^+ + \hat{\beta}_1^+ x_i}} + (1-\hat{p})e^{-e^{\hat{\beta}_0^- + \hat{\beta}_1^- x_i}}}.$$

3. *M–step*: Replace Δ_i with γ_i in equations (3.7 - 3.11) and update the estimates of $\hat{p}$, $\hat{\beta}_0^-$, $\hat{\beta}_1^-$, $\hat{\beta}_0^+$, and $\hat{\beta}_1^+$.

4. Repeat steps 2 and 3 until convergence.

3.1.5 Parameter Estimation

The proposed mixture model of two poisson random variables with the constant mixture probability was implemented using the statistical programming language R. The parameter estimation was implemented via the in-built glm() function of R.

The model was applied to the FDX data in two different ways. As FDX data consists of two different tick sizes throughout the five years, it was decided to separate the data set based on the tick size. First, the model was applied to the two sub sets of FDX data based on the two different tick sizes. Table 3.2 shows the estimates of the FDX data based on the different tick-size.

It can be identified that the mixture probabilities do not show much difference in both the data sets. However, the magnitudes of the slope and intercept parameters of the data set with the bigger tick-size ($\hat{\beta}_0^+, \hat{\beta}_1^+$) are bigger than the the magnitudes of the slope and intercept parameters ($\hat{\beta}_0^-, \hat{\beta}_1^-$), of the smaller tick-size. The signs of the slope parameters ($\hat{\beta}_1^+, \hat{\beta}_1^+$) of both data sets have turned out to be as expected. Further interpretation of the parameters and the model is explained in the section 3.3.

Next the model is applied to each year of the data set separately. Due to the tick-size change in the middle of the fourth year, it was decided to ignore the fourth year. The stock transactions data in the years 1, 2, 3 and 5 were used. The estimates are given in table 3.2. The estimates of the year 5 were smaller than the years 1, 2, and 3. This is not surprising as the tick-size changed from 1/8 to 1/16

	tick-size 1/8	tick-size 1/16
$\hat{\beta}_0^+$	-1.18543053	-0.68856613
$\hat{\beta}_1^+$	4.44756808 e-05	1.09540896 e-05
$\hat{\beta}_0^-$	-1.05153782	-0.51439889
$\hat{\beta}_1^-$	-1.77055672 e-05	-0.99221302 e-05
$\hat{p}$	0.527551	0.5403506

TABLE 3.1

FDX parameter estimation using constant model

Year	$\hat{\beta}_0^+$	$\hat{\beta}_1^+$	$\hat{\beta}_0^-$	$\hat{\beta}_1^-$	$\hat{p}$
1	-1.14039	$4.61744e{-}05$	-0.98357	$-1.42761e{-}05$	0.52359
2	-1.18736	$4.90303e{-}05$	-0.98491	$-2.22884e{-}05$	0.54883
3	-1.16548	$4.68462e{-}05$	-0.82659	$-2.22209e{-}05$	0.56980
5	-0.59867	$1.00522e{-}05$	-0.49124	$-9.16075e{-}06$	0.52991

TABLE 3.2

Parameter estimates for the FDX data using the model with constant mixing probabilities for years 1994, 1995, 1996, and 1998.

during 4th year that made the stock price movements by a smaller value during the year 5.

3.1.6 Simulation

A simulation study was conducted to assess the performance of the estimation methods based on the EM algorithm. Data sets were generated according to the model described under 'Model Formulation' with sizes $10^2, 10^3, 10^4$ and 10^5.

n	$\hat{p}$	$\hat{\beta}_0^+$	$\hat{\beta}_1^+$	$\hat{\beta}_0^-$	$\hat{\beta}_1^-$
True	0.35	-0.5	0.2	-0.7	-0.1
10^2	0.372423779	-0.571396076	0.195167803	-0.687471677	-0.101024391
10^3	0.351164389	-0.504860191	0.200677783	-0.696302985	-0.104609607
10^4	0.350694262	-0.501873677	0.198851347	-0.698943031	-0.099952896
10^5	0.350676824	-0.501917391	0.200009997	-0.699273218	-0.099871496

TABLE 3.3

Simulation : average parameter estimates of 1000 replicates using the constant model

The proposed model with constant mixture probability was applied to each data set and parameters were estimated for 1000 simulated data sets. The average value of each estimate was recorded. Table 3.3 shows the average estimate value of each parameter.

Figures 3.1 to 3.5 show the plot of the estimates of each parameter based on different sizes of the simulated data sets. In each figure the dotted horizontal line denotes the true value of each parameter and the solid dots denote the average estimates.

According to figures 3.1 to 3.5, the average of the estimates are all reasonably close to the true values. Also the average of the estimates generally becomes closer to the true value as the sample size increases. It appears that the modes produce consistent estimates of the parameters.

Estimated mean for $\hat{p}$ based on 1000 replicates
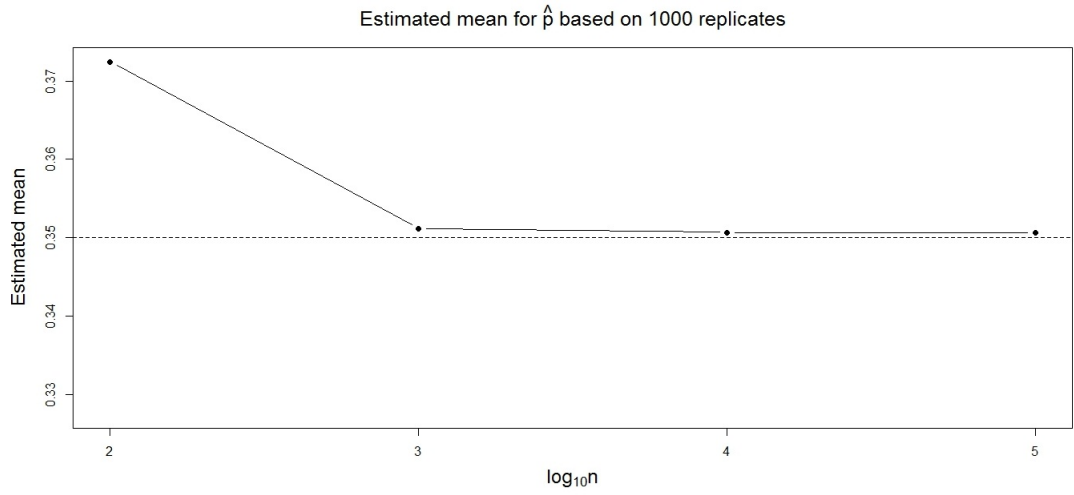
FIGURE 3.1 – Estimated mean for $\hat{p}$ based on 1000 replicates

Estimated mean for $\hat{\beta}_0^+$ based on 1000 replicates
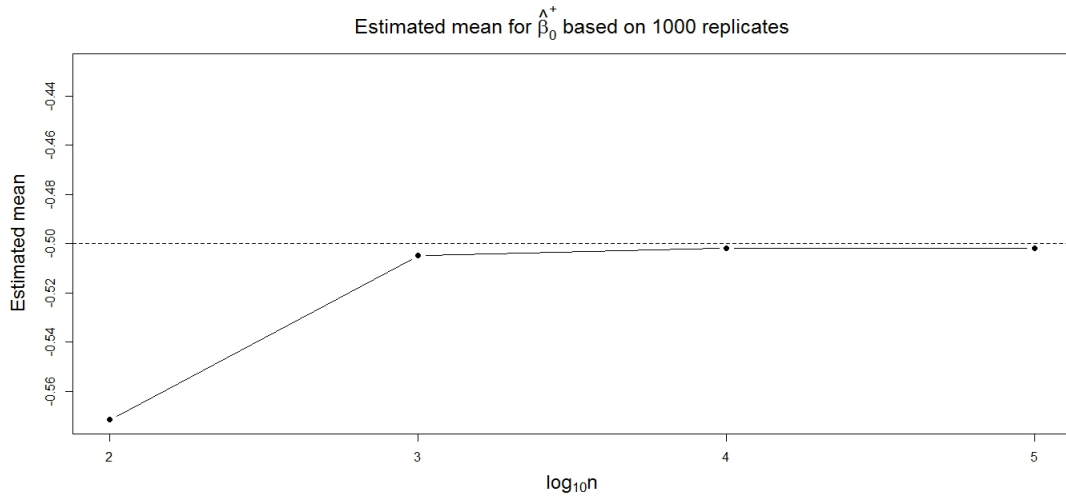
FIGURE 3.2 – Estimated mean for $\hat{\beta}_0^+$ based on 1000 replicates
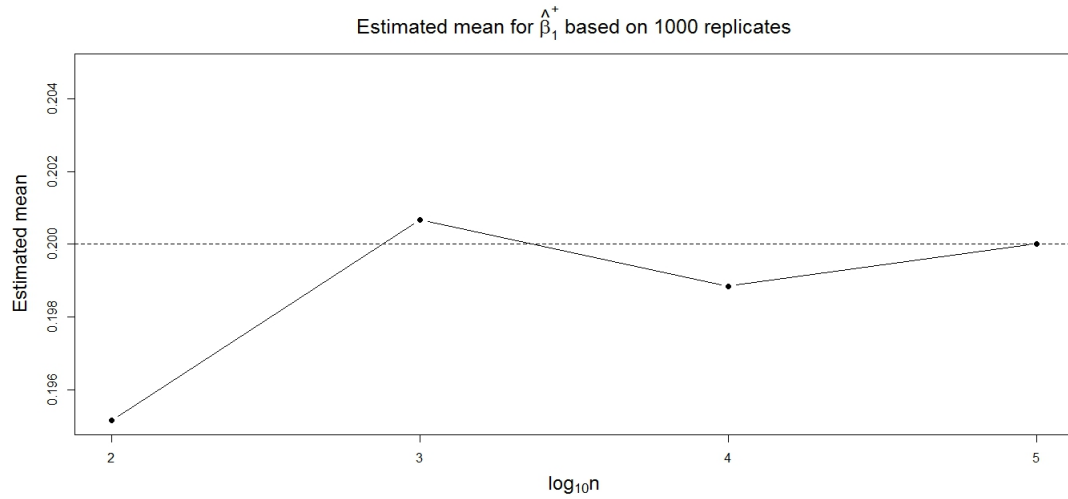
FIGURE 3.3 – Estimated mean for $\hat{\beta}_1^+$ based on 1000 replicates
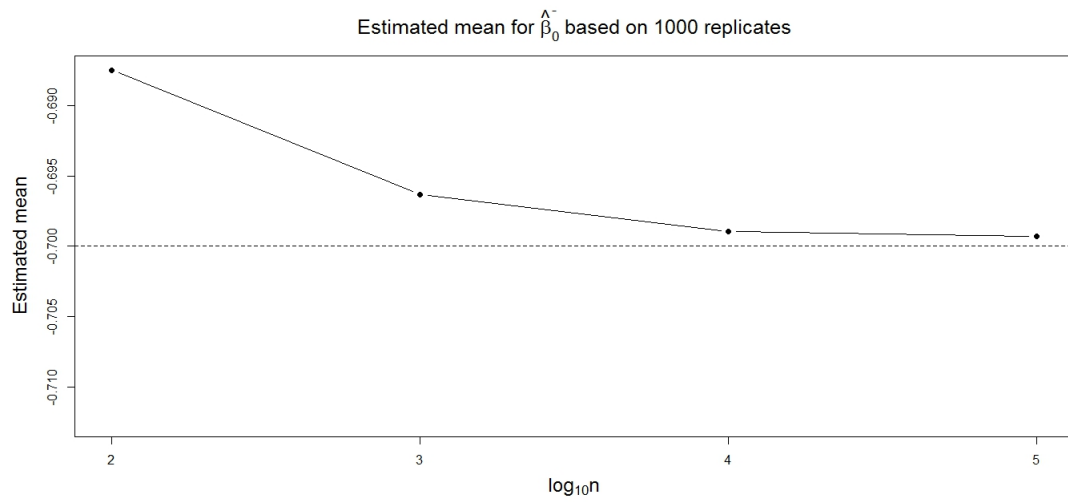
FIGURE 3.4 – Estimated mean for $\hat{\beta}_0^-$ based on 1000 replicates

Estimated mean for $\hat{\beta}_1^-$ based on 1000 replicates
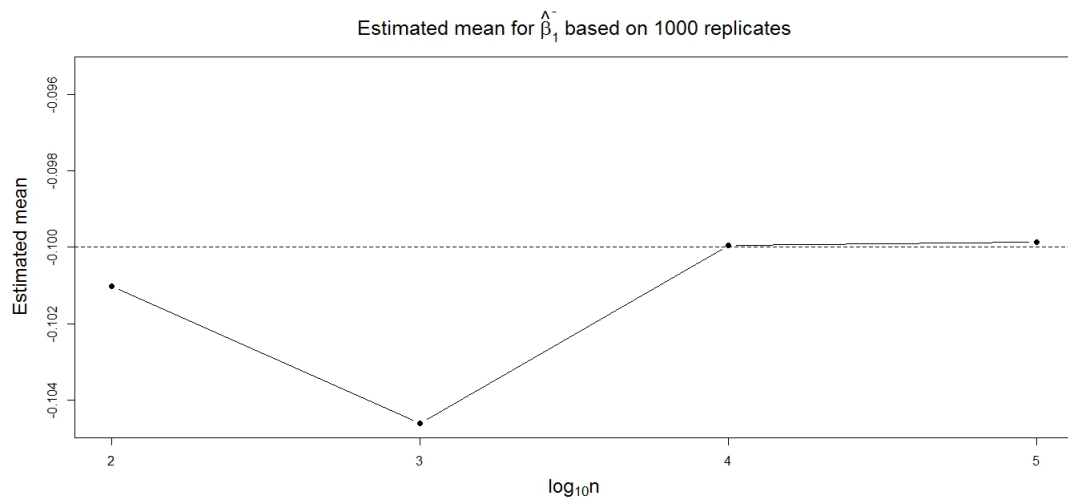
FIGURE 3.5 – Estimated mean for $\hat{\beta}_1^-$ based on 1000 replicates

3.2 Variable Mixing Probabilities

A mixture of two poisson distributions was formulated with the constant mixing probability in the section 3.1. As discussed previously, there is a large effect of the order size towards changing the stock price. Order size was included as the predictor variable in the model. Furthermore, it was decided to check the influence of the order size on the mixture probability. Therefore, the proposed mixture model is extended further to accommodate the mixture probability p_i as a function of order size x_i. Thus, a variable mixture probability is modeled by a 'logistic regression model' in which,

$$p_i = \frac{\exp(\alpha_0 + \alpha_1 x_i)}{1 + \exp(\alpha_0 + \alpha_1 x_i)}. \tag{3.12}$$

When calculating the responsibilities denoted by γ_i in the E-step, the p in the constant model is replace by p_i as shown in (3.12). This results in a complex expression for γ_i when $y_i = 0$ as given in (3.13).

$$
\begin{aligned}
\text{If } y_i = 0, \text{then} \quad \gamma_i &= \frac{\hat{\pi}_i e^{-\hat{\lambda}_i^+}}{\hat{\pi}_i e^{-\hat{\lambda}_i^+} + (1-\hat{\pi}_i)e^{-\hat{\lambda}_i^-}} \\
&= \frac{\frac{e^{\alpha_0+\alpha_1 x_i}}{1+e^{\alpha_0+\alpha_1 x_i}}e^{-e^{\hat{\beta}_0^+ + \hat{\beta}_1^+ x_i}}}{\frac{e^{\alpha_0+\alpha_1 x_i}}{1+e^{\alpha_0+\alpha_1 x_i}}e^{-e^{\hat{\beta}_0^+ + \hat{\beta}_1^+ x_i}} + (1 - \frac{e^{\alpha_0+\alpha_1 x_i}}{1+e^{\alpha_0+\alpha_1 x_i}})e^{-e^{\hat{\beta}_0^- + \hat{\beta}_1^- x_i}}} \\
&= \frac{e^{(\alpha_0+\alpha_1 x_i)}e^{-e^{(\hat{\beta}_0^+ + \hat{\beta}_1^+ x_i)}}}{e^{(\alpha_0+\alpha_1 x_i)}e^{-e^{(\hat{\beta}_0^+ + \hat{\beta}_1^+ x_i)}} + e^{-e^{(\hat{\beta}_0^- + \hat{\beta}_1^- x_i)}}} \\
&= \frac{e^{(\alpha_0+\alpha_1 x_i)}}{e^{(\alpha_0+\alpha_1 x_i)} + e^{-e^{(\hat{\beta}_0^- + \hat{\beta}_1^- x_i)}}e^{e^{(\hat{\beta}_0^+ + \hat{\beta}_1^+ x_i)}}} \\
&= \frac{\exp\{\alpha_0 + \alpha_1 x_i\}}{\exp\{\alpha_0 + \alpha_1 x_i\} + \exp\{e^{(\hat{\beta}_0^+ + \hat{\beta}_1^+ x_i)} - e^{(\hat{\beta}_0^- + \hat{\beta}_1^- x_i)}\}} \tag{3.13}
\end{aligned}
$$

The complete log-likelihood is similar to the expression given in the equation (3.6) except that with the variable mixing probabilities p will be replaced by p_i as given in the equation (3.12).

M - Step

$$l_0(\alpha_0, \alpha_1, \beta_0^-, \beta_1^-, \beta_0^+, \beta_1^+) = \sum_{i=1}^n (1 - \Delta_i) \ln P(y_i; \beta_0^-, \beta_1^-) + \sum_{i=1}^n \Delta_i \ln P(y_i; \beta_0^+, \beta_1^+)$$
$$+ \sum_{i=1}^n (1 - \Delta_i) \ln(1 - \pi_i) + \sum_{i=1}^n \Delta_i \ln \pi_i$$

$$l_0(\alpha_0, \alpha_1, \beta_0^-, \beta_1^-, \beta_0^+, \beta_1^+) = \sum_{i=1}^n (1 - \Delta_i) \ln P(y_i; \beta_0^-, \beta_1^-) + \sum_{i=1}^n \Delta_i \ln P(y_i; \beta_0^+, \beta_1^+)$$
$$+ \sum_{i=1}^n (1 - \Delta_i) \ln(1 - \frac{e^{\alpha_0 + \alpha_1 x_i}}{1 + e^{\alpha_0 + \alpha_1 x_i}})$$
$$+ \sum_{i=1}^n \Delta_i \ln(\frac{e^{\alpha_0 + \alpha_1 x_i}}{1 + e^{\alpha_0 + \alpha_1 x_i}})$$

$$l_0(\alpha_0, \alpha_1, \beta_0^-, \beta_1^-, \beta_0^+, \beta_1^+) = \sum_{i=1}^n (1 - \Delta_i) \ln P(y_i; \beta_0^-, \beta_1^-) + \sum_{i=1}^n \Delta_i \ln P(y_i; \beta_0^+, \beta_1^+)$$
$$- \sum_{i=1}^n (1 - \Delta_i) \ln(1 + e^{\alpha_0 + \alpha_1 x_i})$$
$$- \sum_{i=1}^n \Delta_i \ln(1 + e^{\alpha_0 + \alpha_1 x_i})$$
$$+ \sum_{i=1}^n \Delta_i(\alpha_0 + \alpha_1 x_i)$$

$$l_0(\alpha_0, \alpha_1, \beta_0^-, \beta_1^-, \beta_0^+, \beta_1^+) = \sum_{i=1}^n (1 - \Delta_i) \ln P(y_i; \beta_0^-, \beta_1^-) + \sum_{i=1}^n \Delta_i \ln P(y_i; \beta_0^+, \beta_1^+)$$
$$- \sum_{i=1}^n \ln(1 + e^{\alpha_0 + \alpha_1 x_i}) + \sum_{i=1}^n \Delta_i(\alpha_0 + \alpha_1 x_i) \qquad (3.14)$$

Similarly the complete log-likelihood of the variable model is differentiated with respect to each parameter and maximized to obtain the estimates. When extending the initial model with variable mixing probabilities, only p was changed to p_i. Therefore, instead of differentiating with respect to p now there are two new parameters in the model; α_0 and α_1 replace p in the new model.

Finding $\hat{\alpha}_0$:

$$\frac{\partial l_0}{\partial \alpha_0} = \sum_{i=1}^{n} \Delta_i - \sum_{i=1}^{n} \frac{e^{\alpha_0 + \alpha_1 x_i}}{1 + e^{\alpha_0 + \alpha_1 x_i}}$$

$$\Rightarrow \sum_{i=1}^{n} \Delta_i - \sum_{i=1}^{n} \frac{e^{\alpha_0 + \alpha_1 x_i}}{1 + e^{\alpha_0 + \alpha_1 x_i}} = 0$$

$$\Rightarrow \sum_{i=1}^{n} \{\Delta_i - \frac{e^{\alpha_0 + \alpha_1 x_i}}{1 + e^{\alpha_0 + \alpha_1 x_i}}\} = 0 \qquad (3.15)$$

Finding $\hat{\alpha}_1$:

$$\frac{\partial l_0}{\partial \alpha_1} = \sum_{i=1}^{n} \Delta_i x_i - \sum_{i=1}^{n} \frac{e^{\alpha_0 + \alpha_1 x_i}}{1 + e^{\alpha_0 + \alpha_1 x_i}} x_i$$

$$\Rightarrow \sum_{i=1}^{n} \Delta_i x_i - \sum_{i=1}^{n} \frac{e^{\alpha_0 + \alpha_1 x_i}}{1 + e^{\alpha_0 + \alpha_1 x_i}} x_i = 0$$

$$\Rightarrow \sum_{i=1}^{n} \{\Delta_i - \frac{e^{\alpha_0 + \alpha_1 x_i}}{1 + e^{\alpha_0 + \alpha_1 x_i}}\} x_i = 0 \qquad (3.16)$$

After simplifications the score equations for the model with variable mixing probabilities consist of the equations similar to the equations (3.8) to (3.11) of the first model and two new equations (3.15) and (3.16).

$$\sum_{i=1}^{n} (1 - \Delta_i)\{y_i - e^{\beta_0^- + \beta_1^- x_i}\} = 0 \qquad (3.8)$$

$$\sum_{i=1}^{n} (1 - \Delta_i)\{y_i - e^{\beta_0^- + \beta_1^- x_i}\} x_i = 0 \qquad (3.9)$$

$$\sum_{i=1}^{n} \Delta_i \{y_i - e^{\beta_0^+ + \beta_1^+ x_i}\} = 0 \qquad (3.10)$$

$$\sum_{i=1}^{n} \Delta_i \{y_i - e^{\beta_0^+ + \beta_1^+ x_i}\} x_i = 0 \qquad (3.11)$$

$$\sum_{i=1}^{n} \{\Delta_i - \frac{e^{\alpha_0 + \alpha_1 x_i}}{1 + e^{\alpha_0 + \alpha_1 x_i}}\} = 0 \qquad (3.15)$$

$$\sum_{i=1}^{n} \{\Delta_i - \frac{e^{\alpha_0 + \alpha_1 x_i}}{1 + e^{\alpha_0 + \alpha_1 x_i}}\} x_i = 0 \qquad (3.16)$$

If Δ_i's are known, then the estimates of β_0^-, β_1^-, β_0^+, and β_1^+ can be obtained as they were in the model with constant mixing probabilities, and the estimates of

α_0 and α_1 can be obtained from (3.15) and (3.16) using a logistic regression. Although each Δ_i is either 0 or 1, the EM algorithm will replace it with its expected value γ_i which might not be an endpoint of $[0, 1]$. Consequently, the implementation of the method for computing estimates for logistic regression must allow for fractional response values.

The EM algorithm for the model with variable mixing probabilities then proceeds as follows.

1. Take initial guesses $\hat{\alpha}_0$, $\hat{\alpha}_1$, $\hat{\beta}_0^-$, $\hat{\beta}_1^-$, $\hat{\beta}_0^+$, and $\hat{\beta}_1^+$ for the parameters.

2. *E–step*: Compute the expected values of each Δ_i using (γ_i). The calculations are similar to the constant case except that, if $y_i = 0$, then

$$
\begin{aligned}
\gamma_i &= \frac{\hat{p}_i e^{-\hat{\lambda}_i^+}}{\hat{p}_i e^{-\hat{\lambda}_i^+} + (1 - \hat{p}_i)e^{-\hat{\lambda}_i^-}} \\
&= \frac{\exp\left(\hat{\alpha}_0 + \hat{\alpha}_1 x_i\right)}{\exp\left(\hat{\alpha}_0 + \hat{\alpha}_1 x_i\right) + \exp(e^{\hat{\beta}_0^+ + \hat{\beta}_1^+ x_i} - e^{\hat{\beta}_0^- + \hat{\beta}_1^- x_i})}.
\end{aligned}
$$

3. *M–step*: Replace Δ_i with γ_i in equations (3.15), (3.16), and (3.8) - (3.11) and update the estimates of $\hat{\alpha}_0$, $\hat{\alpha}_1$, $\hat{\beta}_0^-$, $\hat{\beta}_1^-$, $\hat{\beta}_0^+$, and $\hat{\beta}_1^+$.

4. Repeat steps 2 and 3 until convergence.

3.2.1 Parameter Estimation

Similar to the constant model, the variable model is also applied to the FDX data sets based on different tick-sizes and for different years separately. Table 3.4 shows the parameter estimates of the data sets based on the tick-size.

In the variable model there are two parameters $(\hat{\alpha}_0, \hat{\alpha}_1)$ which are used for the mixture probability. In both the data sets, $\hat{\alpha}_1$ has turned out to be a positive value as expected. Similar to the constant model, the magnitudes of the slope and intercept parameters of the data set with the bigger tick-size $(\hat{\beta}_0^+, \hat{\beta}_1^+)$ are bigger than the the magnitudes of the slope and intercept parameters $(\hat{\beta}_0^-, \hat{\beta}_1^-)$, of the

	tick-size 1/8	tick-size 1/16
$\hat{\beta}_0^+$	-1.192732	-0.7306984
$\hat{\beta}_1^+$	2.678521e-05	1.610377e-05
$\hat{\beta}_0^-$	-1.048063	-0.4843464
$\hat{\beta}_1^-$	-1.4159e-05	-7.095849e-06
$\hat{\alpha}_0$	0.1437902	0.2203968
$\hat{\alpha}_1$	0.005914745	0.002465908

TABLE 3.4

FDX parameter estimation using variable model

smaller tick-size. The signs of the slope parameters $(\hat{\beta}_1^+, \hat{\beta}_1^+)$ of both data sets have turned out to be as expected. Further interpretation of the parameters and the model is explained in section 3.3.

The variable model is applied to each year of the data set separately. The estimates of the years 1, 2, 3 and 5 are given in table 3.5.

3.2.2 Simulation

Similar to the constant model, a simulation study was conducted to assess the performance of the estimation methods based on the EM algorithm. Data sets were generated according to the model described under 'Model Formulation' and with sizes $10^2, 10^3, 10^4$ and 10^5. The proposed model with constant mixture probability was applied to each data set and parameters were estimated as 1000 replicates. The average value of each estimate was recorded. Table 3.6 shows the average estimate value of each parameter.

The figures 3.6 to 3.11 show the plot of the estimates of each parameter based on different sizes of the simulated data sets. In each figure the dotted

Year	$\hat{\beta}_0^+$	$\hat{\beta}_1^+$	$\hat{\beta}_0^-$	$\hat{\beta}_1^-$	$\hat{\alpha}_0$	$\hat{\alpha}_1$
1	-1.09231	$2.42998e{-}05$	-1.03044	$-1.13407e{-}05$	-0.00889	0.00291
2	-1.13954	$1.67870e{-}05$	-1.02011	$-1.87965e{-}05$	0.12750	0.00566
3	-1.12029	$2.53253e{-}05$	-0.88099	$-1.66876e{-}05$	0.29315	0.00702
5	-0.63951	$5.43784e{-}06$	-0.44293	$-6.17989e{-}06$	0.18239	0.00222

TABLE 3.5

Parameter estimates for the FDX data using the model with variable mixing probabilities for years 1994, 1995, 1996, and 1998.

n	$\hat{\beta}_0^+$	$\hat{\beta}_1^+$	$\hat{\beta}_0^-$	$\hat{\beta}_1^-$	$\hat{\alpha}_0$	$\hat{\alpha}_1$
True	-0.5	0.2	-0.7	-0.1	0.3	0.8
10^2	-0.49526	0.17393	-0.67327	-0.07445	0.35865	0.93145
10^3	-0.49681	0.19469	-0.69465	-0.09166	0.30484	0.81542
10^4	-0.49904	0.19919	-0.70192	-0.10158	0.29616	0.80042
10^5	-0.49857	0.19997	-0.70138	-0.10079	0.29688	0.79843

TABLE 3.6

simulation parameter for variable model

Estimated mean for $\hat{\alpha}_0$ based on 1000 replicates
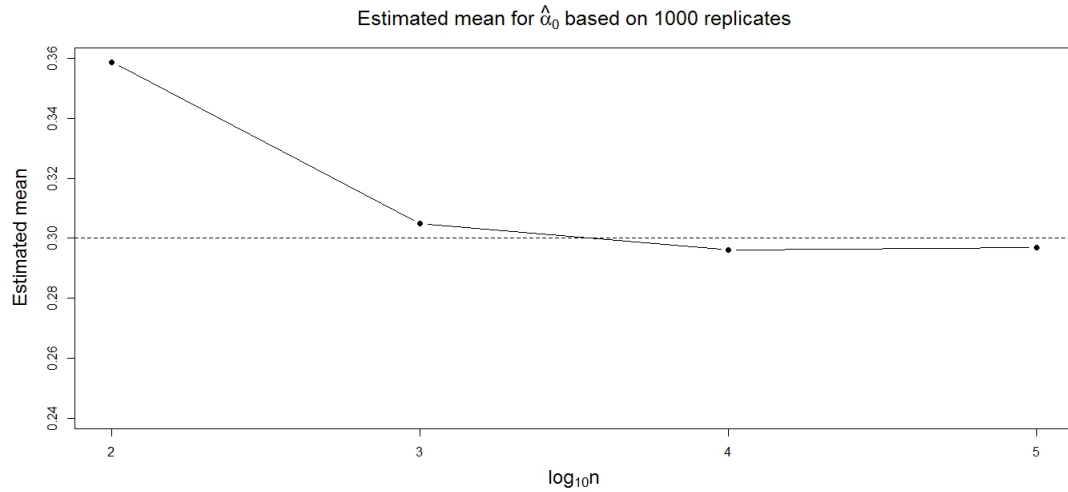
FIGURE 3.6 – Estimated mean for $\hat{\alpha}_0$ based on 1000 replicates

horizontal line denotes the true value of each parameter and the solid dots denote the average estimates. According to figures 3.6 to 3.11, the average of the estimates are all reasonably close to the true values. Also the average of the estimates generally becomes closer to the true value as the sample size increases. It appears that the models produce consistent estimates of the parameters.
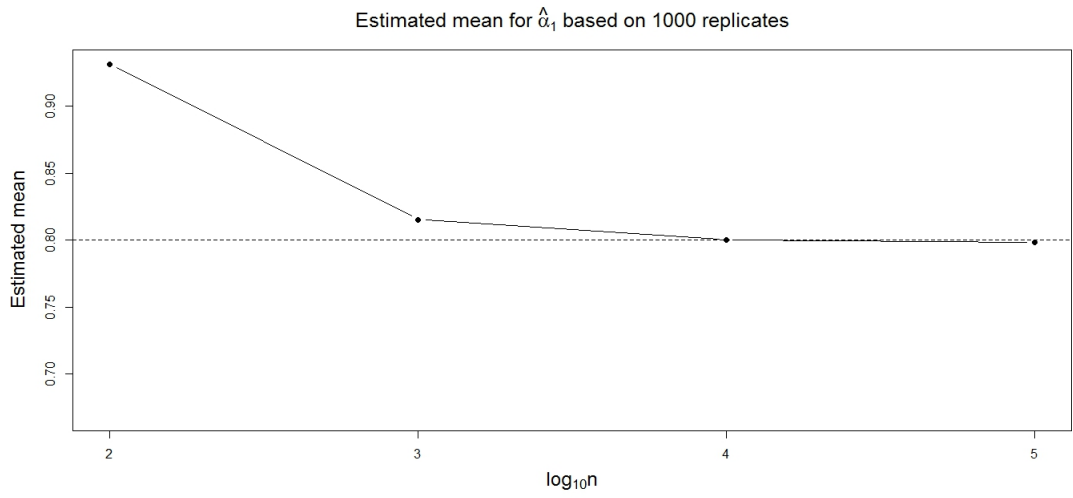
FIGURE 3.7 – Estimated mean for $\hat{\alpha}_1$ based on 1000 replicates
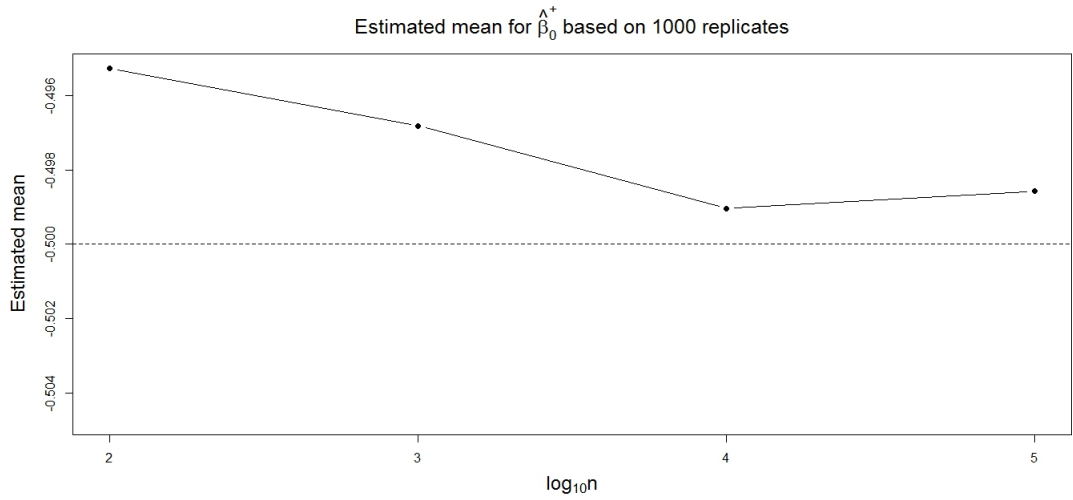
FIGURE 3.8 – Estimated mean for $\hat{\beta}_0^+$ based on 1000 replicates
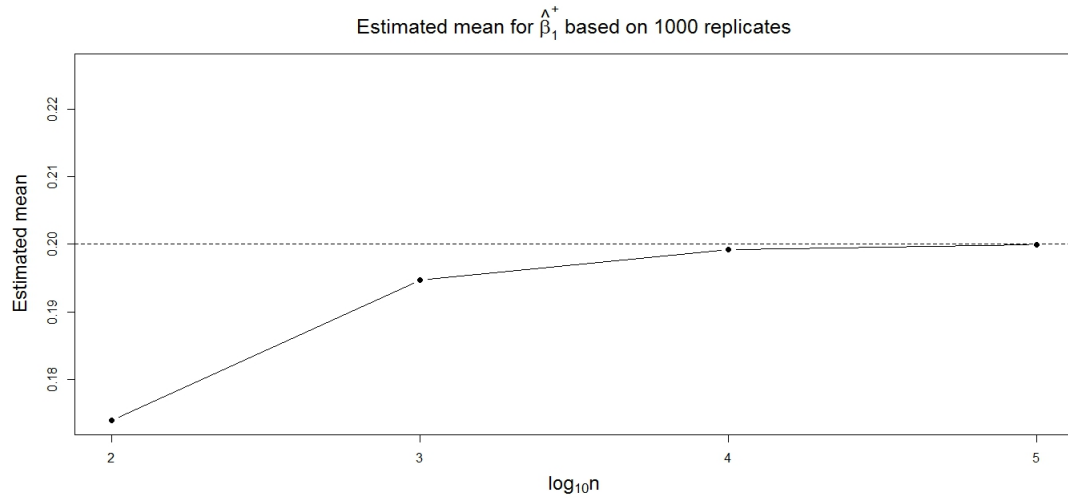
FIGURE 3.9 – Estimated mean for $\hat{\beta}_1^+$ based on 1000 replicates
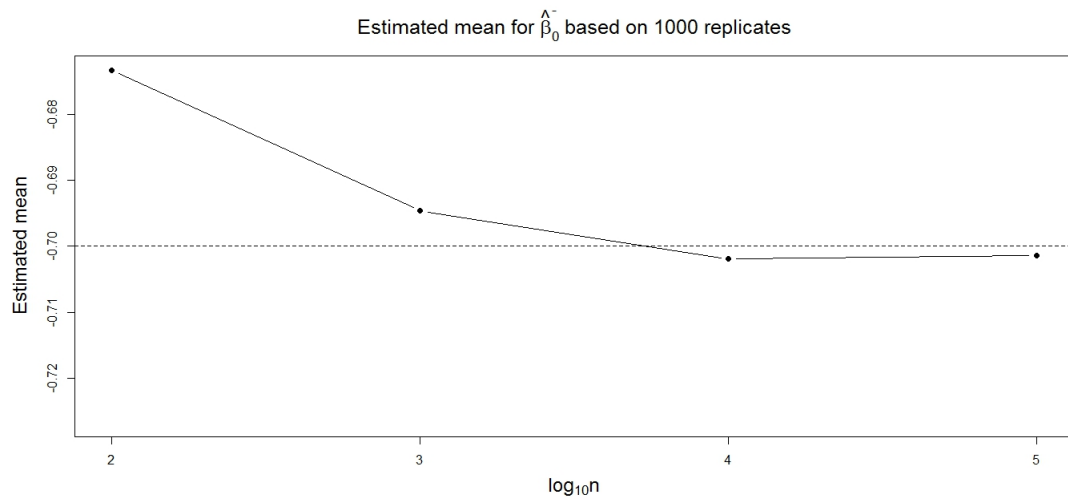
FIGURE 3.10 – Estimated mean for $\hat{\beta}_0^-$ based on 1000 replicates

Estimated mean for $\hat{\beta}_1^-$ based on 1000 replicates
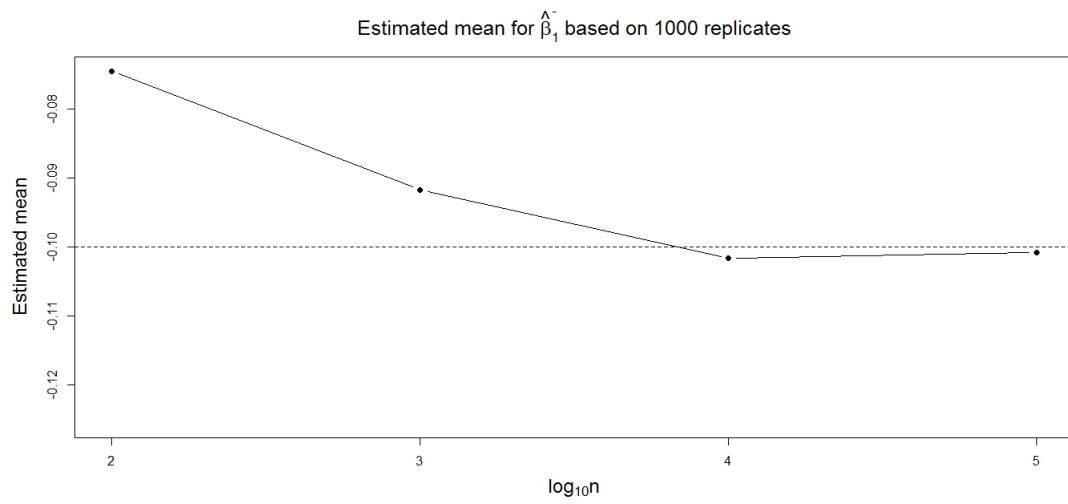
FIGURE 3.11 – Estimated mean for $\hat{\beta}_1^-$ based on 1000 replicates

3.3 Interpretation

In sections 3.1 and 3.2 the stock price increments and decrements were modeled using the concepts 'Poisson Regression' and 'Mixture Models'. The parameters were estimated using the 'Method of Maximum Likelihood'. This section explains the interpretation of the estimated parameters.

Unlike in simple linear regression where the response follows a normal distribution with an identity link function, interpretation of poisson regression is not straight forward. Due to the choice of logarithm link function in the poisson regression, the unit change can not be linearly expressed. Therefore, interpretation in terms of a relative change of the mean would simplify the complexity.

In calculating relative change of the mean of the price change, it is important to consider that the orders are placed by multiples of hundreds. Therefore, a change in a single unit means order size changing by one hundred.

The relative change of the mean of the price change is given by

$$\frac{\lambda_{x \pm 100}}{\lambda_x} = \frac{e^{\beta_0 + \beta_1(x \pm 100)}}{e^{\beta_0 + \beta_1 x}} = e^{\pm 100\beta_1}. \qquad (3.17)$$

Using expression (3.17), the relative increment of the stock price change is calculated using $e^{100\beta_1^+}$ and the relative decrement of the stock price change is calculated by $e^{-100\beta_1^-}$. It uses only the slope parameter.

3.3.1 Interpretation of the Parameters

The stock price change (both increment and decrement) was modeled using mixture model where each sub population is modeled using poisson regression with the use of logarithm link function. Each sub population can be expressed using the log-linear model.

The stock price increment can be expressed using a log linear model with

average stock price increment λ^+ and the order size x_i.

$$\log(\lambda^+) = \hat{\beta}_0^+ + \hat{\beta}_1^+ x_i \qquad (3.18)$$

The estimates of the FDX data set can be used to further describe the expression given in (3.18). For example, the slope and intercept parameters of the FDX data set with 1/8 tick-size was estimated as $\hat{\beta}_0^+ = $ -1.18 and $\hat{\beta}_1^+ = 0.00004$. The the log-linear model for average stock price increment when the tick-size for 1/8 is given by : log(average stock price increment)=-1.18 + $0.00004x_i$.

The log-linear model of the average stock price increment has a positive slope. With the positive slope, the logarithm of the average stock price increment increases as the order size increases. This reflects the fact that the average log stock price increments increase with the increasing order size. This conforms the stock market behavior that, when more stocks are purchased, the price of a stock will increase more.

Similarly, stock price decrement can be expressed using a log linear model with average stock price decrement λ^- and the order size x_i.

$$\log(\lambda^-) = \hat{\beta}_0^- + \hat{\beta}_1^- x_i \qquad (3.19)$$

For example, the slope and intercept parameters of the FDX data set of stock price decrements with 1/8 tick-size was estimated as $\hat{\beta}_0^- = $ -1.05 and $\hat{\beta}_1^- = $ - 0.000018. The the log-linear model for average stock price increment when the tick-size for 1/8 is given by : log(average stock price decrement)=-1.05 - $0.000018x_i$.

The log-linear model of the average stock price decrements has a negative slope. With the negative slope, the logarithm of the average stock price increment decreases as the order size increases. This reflects the fact that the average log stock price decrements decrease with the increasing order size. This conforms the stock market behavior that, when more stocks are sold, the price of a stock will decrease more.

3.3.2 Probability of Stock Price Change

It is also interesting to find probabilities of discrete stock price changes based on the estimates of the parameters. The probabilities of the discrete stock price changes are calculated as given in (3.21), (3.22) and (3.23).

$P(y_i > 0)$ denotes the probability of discrete stock price increment, $P(y_i < 0)$ denotes the probability of discrete stock price decrement and $P(y_i = 0)$ denotes the probability that the stock price stays same between two consecutive transactions.

The probability of discrete stock price increment is calculated by :

$$
\begin{aligned}
P(Y_i > 0) &= P(\Delta_i = 1 \text{ and } Y_i > 0) \\
&= P(\Delta_i = 1)P(Y_i > 0|\Delta_i = 1) \\
&= p_i P(Y_i^+ > 0) \\
&= p_i[1 - P(Y_i^+ = 0)] \\
P(Y_i > 0) &= p_i(1 - e^{-\lambda_i^+})
\end{aligned}
\tag{3.20}
$$

The probability of discrete stock price decrement is calculated by :

$$
\begin{aligned}
P(Y_i < 0) &= P(\Delta_i = 0 \text{ and } Y_i < 0) \\
&= P(\Delta_i = 0)P(Y_i < 0|\Delta_i = 0) \\
&= (1 - p_i)P(Y_i^- < 0) \\
&= (1 - p_i)[1 - P(Y_i^- = 0)] \\
P(Y_i < 0) &= (1 - p_i)(1 - e^{-\lambda_i^-})
\end{aligned}
\tag{3.21}
$$

The probability that the stock price stays same :

$$
\begin{aligned}
P(Y_i = 0) &= P(\Delta_i = 0 \text{ and } Y_i = 0) + P(\Delta_i = 1 \text{ and } Y_i = 0) \\
&= P(\Delta_i = 0)P(Y_i = 0|\Delta_i = 0) + P(\Delta_i = 1)P(Y_i = 0|\Delta_i = 1) \\
&= (1 - p_i)P(Y_i^- = 0) + p_i P(Y_i^+ = 0) \\
P(Y_i = 0) &= (1 - p_i)e^{-\lambda_i^-} + p_i e^{-\lambda_i^+} \tag{3.22}
\end{aligned}
$$

The probabilities given by expressions (3.21), (3.22) and (3.23), $p_i = \frac{e^{\alpha_0 + \alpha_1 x_i}}{1 + e^{\alpha_0 + \alpha_1 x_i}}$, $\lambda_i^- = e^{\beta_0^- + \beta_1^- x_i}$ and $\lambda_i^+ = e^{\beta_0^+ + \beta_1^+ x_i}$. The actual probabilities can be calculated using the estimates of the parameters in the FDX dataset for given order size (x_i). The figures 3.12, 3.13, 3.14 and 3.15 show the probabilities of the discrete stock price changes based on different order sizes. The estimates are based on table 3.5.

According to figures 3.12, 3.13 and 3.14, the probability of no price change between two consecutive transactions with respect to the order size, $P(Y_i = 0)$, is between 0.5 and 0.6 for the first three years. In the same three years, the probabilities of stock price increment($P(Y_i > 0)$) and decrement ($P(Y_i < 0)$) between two consecutive transactions on order size are below 0.4. These are the first three years of the FDX dataset with same tick size 1/8.

Once the tick size is changed during the fourth year, the behavior in fifth year is different from the first three years. During the fifth year (figure 3.15) the probabilities of no price change between two consecutive transactions ($P(Y_i = 0)$) has decreased to be below 0.5. However the upper limits of the probabilities of stock price increment($P(Y_i > 0)$) and decrement ($P(Y_i < 0)$) between two consecutive transactions on order size, has increased to 0.45. This stock price volatility can be justified with the smaller tick size as the minimum possible change (tick-size) is smaller price tends to move faster.

These facts on the probabilities $P(Y_i = 0)$, $P(Y_i > 0)$ and $P(Y_i < 0)$ in figures 3.12, 3.13, 3.14 and 3.15, further conform the model with the expectations of the stock market.
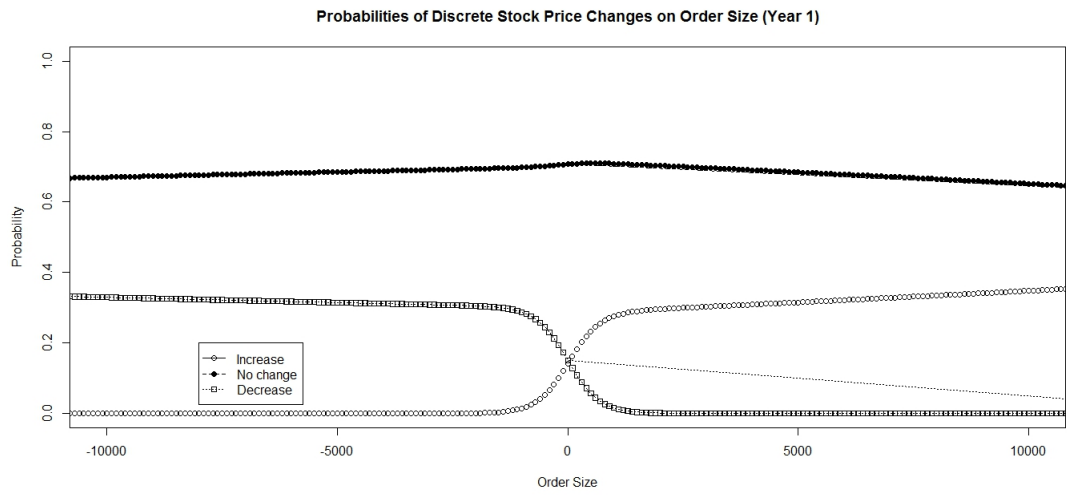
FIGURE 3.12 – Probabilities of discrete stock price changing on order size in year 1
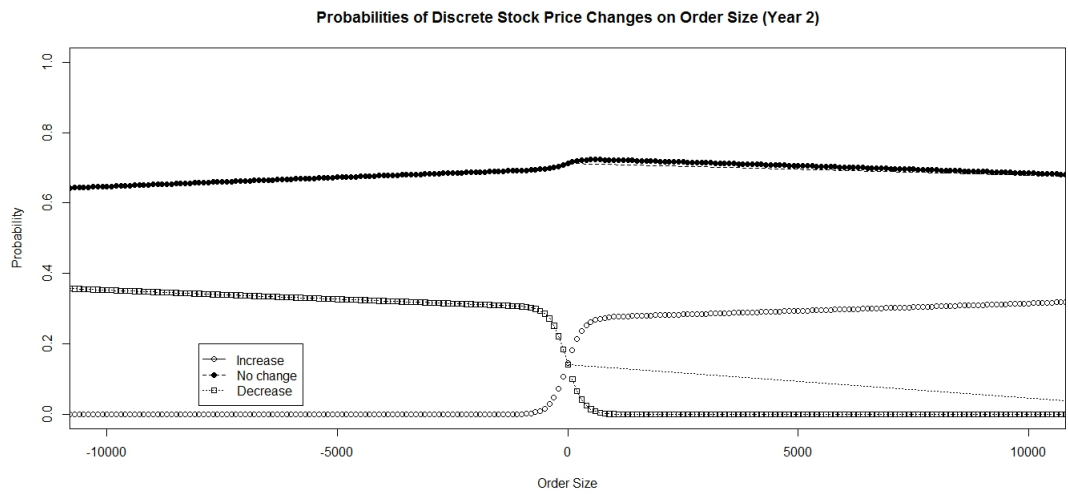
FIGURE 3.13 – Probabilities of discrete stock price changing on order size in year 2
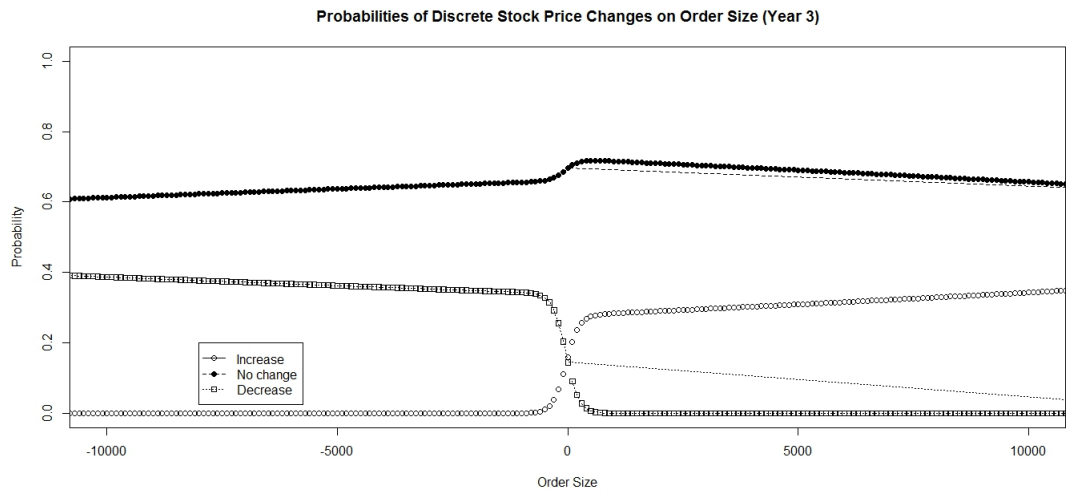
FIGURE 3.14 – Probabilities of discrete stock price changing on order size in year 3
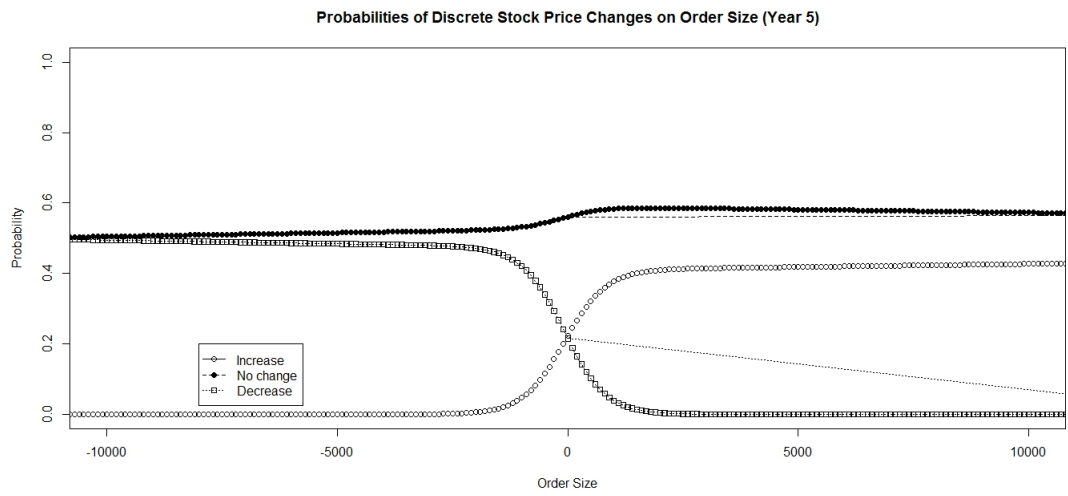
FIGURE 3.15 – Probabilities of discrete stock price changing on order size in year 5

CHAPTER 4
EFFICIENCY IMPROVEMENTS

The proposed mixture of poisson distributions is implemented using the statistical programming language **R**. R is used as a popular statistical software package. There are several built-in statistical functions that facilitate carrying out statistical modeling.

In estimating the parameters of the model, the in-built function $glm()$ of R is used. However, the execution times of both models in R were not satisfactory. Therefore, the code was further investigated to identify the possibilities of reducing the execution time.

4.1   Improvements in the Code

During the initial implementation of the model, the $glm()$ function of R was used when estimating the slope (β_1^-) and intercept (β_0^-) parameters of each model. However, the $glm()$ function in R does not only output the intercept and the slope parameter. It outputs the values for parameters such as Akaike Information Criterion (AIC), degrees of freedom and residual deviance in addition. Therefore, this adds a heavier work within the designed algorithms, thus taking more time to provide the required output. EM algorithm is usually known to be slow. When $glm()$ function in R is used with EM algorithm the execution time gets bigger than expected.

Therefore, the first task in improving the efficiency was to replace $glm()$

Constant Mixing Probability				
	using $glm()$		using NR method	
Size	Time	Iterations	Time	Iterations
100	0.25	33	0.05	33
1000	1.07	45	0.16	45
10000	9.72	43	1.01	43
100000	94.61	42	10.08	42
1000000	967.69	43	105.83	43

TABLE 4.1

Comparison of Efficiency with the user written NR method in the constant case

function with a function that does only what is required for the execution of the proposed model. As a solution the 'Newton Raphson' (NR) algorithm was implemented in R. This helped in reducing the execution time. The Newton Raphson method described in Garthwaite *et al.* (2002, pp 44-45) is used for the above task.

Tables 4.1 and 4.2 show the comparison of execution time and the number of iterations of using NR method and the $glm()$ function in both models with different sizes of simulation data sets. Figure 4.1 - 4.4 further summarizes the data in tables 4.1 and 4.2. In figures 4.1 and 4.3, it is evident that use of the NR method has reduced the execution time greatly. Figure 4.5 shows the time difference between the two methods. It is clear that use of the NR method instead of the built-in $glm()$ function has increased the efficiency of the execution time. According to figure 4.5 the amount of the time saved will increase exponentially as the size of the data set increase.

However, the figures do not provide a consistent evidence on the number of iterations.

Variable Mixing Probability				
	using $glm()$		using NR method	
Size	Time	Iterations	Time	Iterations
100	0.49	51	0.08	27
1000	0.97	31	0.18	32
10000	9.36	31	1.18	38
100000	101.03	35	10.36	32
1000000	910.83	32	111.64	32

TABLE 4.2

Comparison of Efficiency with the user written NR method in the Variable Case
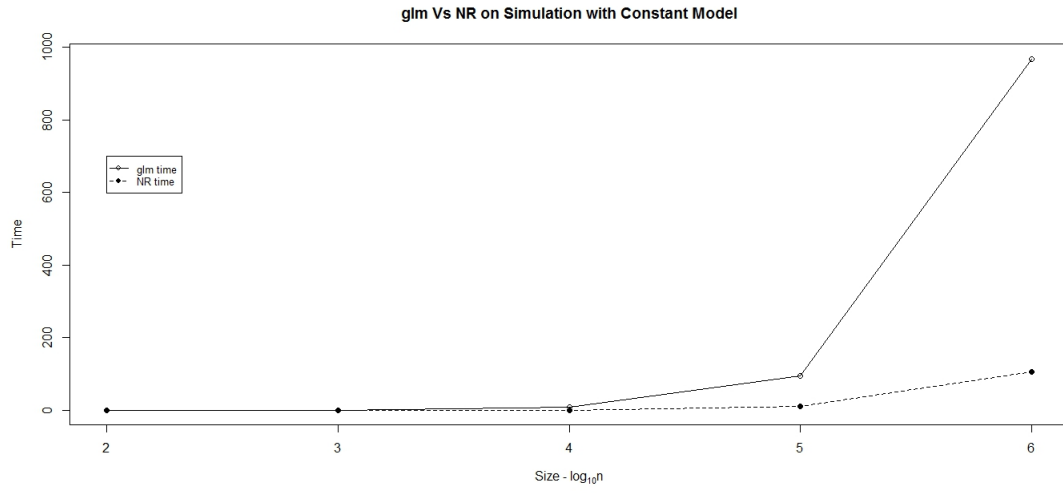
FIGURE 4.1 – Time of $glm()$ vs Newton Raphson for constant model with simulation. Solid line denotes the glm function and the dotted line denotes the NR method
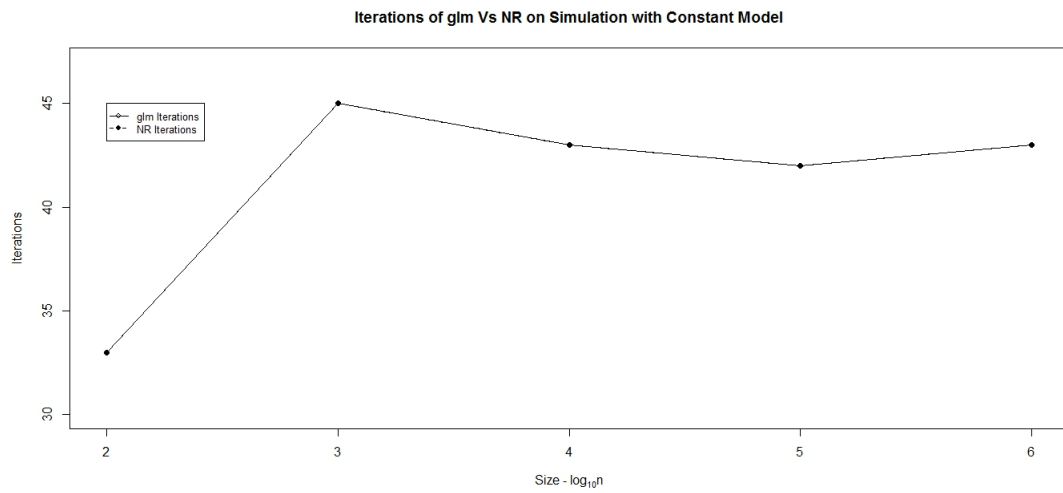
FIGURE 4.2 – Iterations of $glm()$ vs Newton Raphson for constant model with simulation
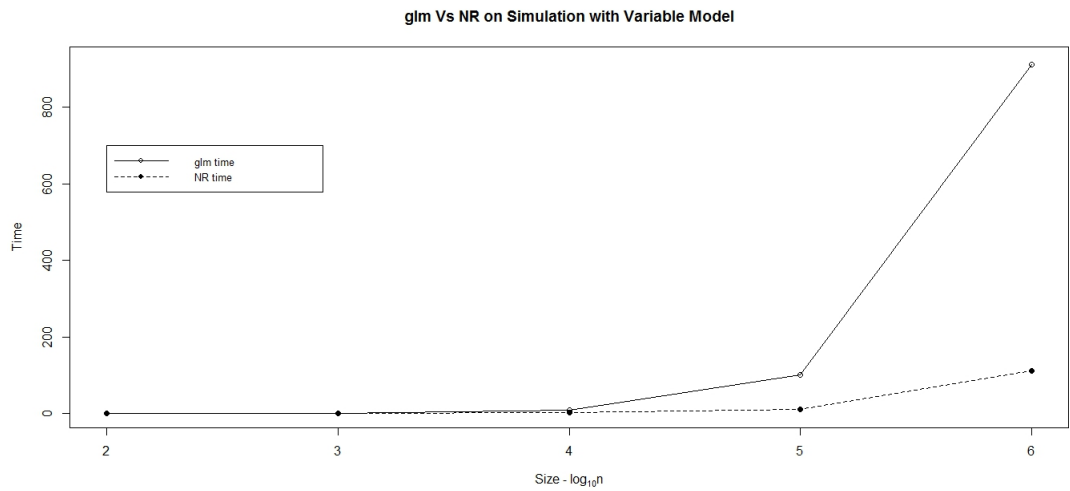
FIGURE 4.3 – Time of $glm()$ vs Newton Raphson for Variable model with simulation
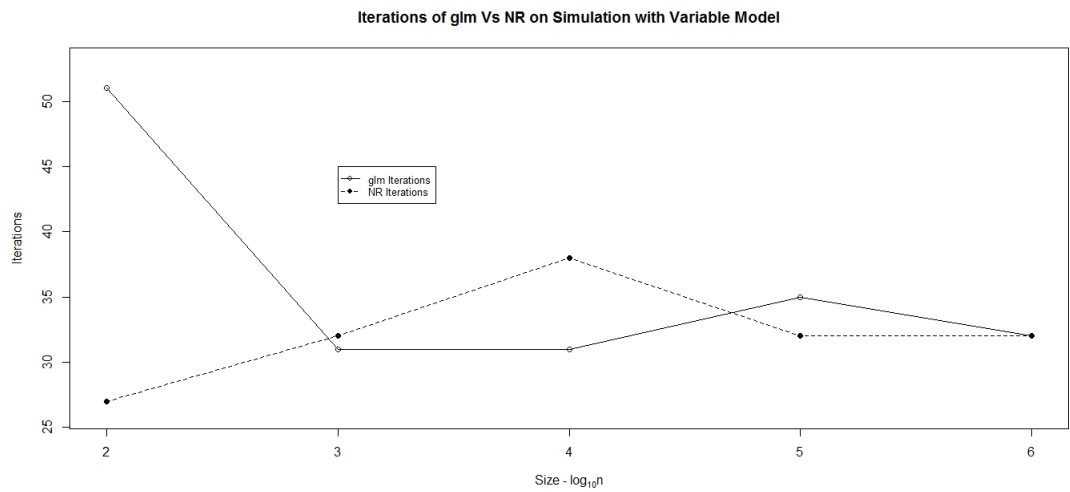
FIGURE 4.4 – Iterations of $glm()$ vs Newton Raphson for Variable model with simulation
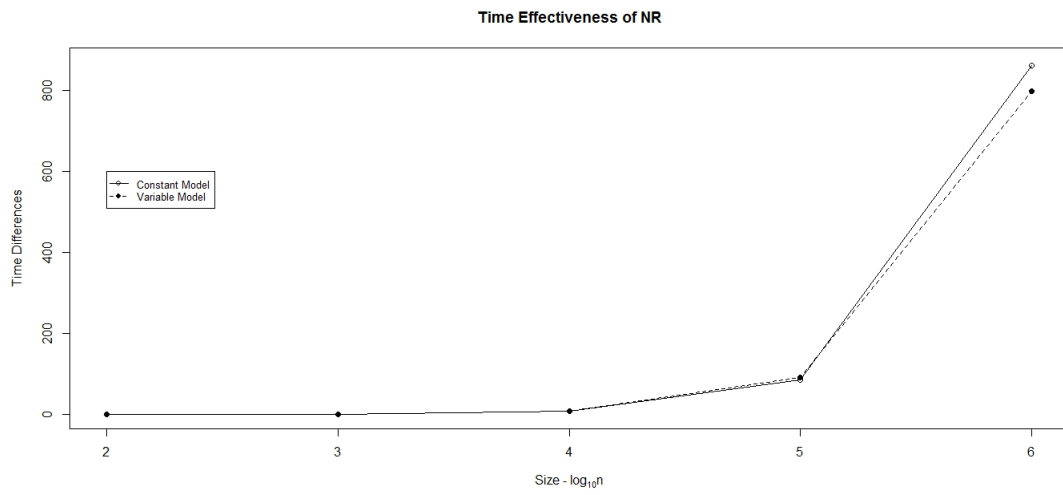
FIGURE 4.5 – Time Effectiveness of NR method in both models

4.2 Parabolic EM Algorithm

As outlined in section 1.4, there have been a significantly large amount of discussions in improving the efficiency of the EM algorithm which was introduced in 1977. This section implements one of the most recent and relevant improvements of the EM algorithm introduced by Berlinet and Roland in 2009 (Berlinet and Roland, 2009). The algorithm is named the 'Parabolic EM' (PEM) and uses the concept of the 'Bazier Parabola'. As highlighted by the authors, the implementation of PEM on mixture models of two poisson distributions has exhibited a significantly larger acceleration by a factor of 22, with no failures (Berlinet and Roland, 2012).

Berlinet and Roland (2012) have demonstrated that the effectiveness of PEM by comparing recent acceleration algorithms based on the behavior and theoretical formulation. Among several recent accelerations, the use of PEM in a mixture of poisson distributions shows its relevance to the proposed model. The authors have proved that "the sequences generated by PEM do not decrease the likelihood".

Therefore, it was decided to investigate using PEM in the implementation of the proposed model. The next section presents the basic idea of the PEM algorithm, as presented by the original authors in Berlinet and Roland (2012).

4.2.1   PEM Algorithm

According to Berlinet and Roland (2012), the PEM is designed based on the concept of the 'Bezier Parabola'. It uses three initial points, which are called 'control points', to control the arc of the parabola. These three control points form a triangle, known as a control triangle, containing the arc of the parabola. Under the properties of the Bezier parabola, the $n + 1$ number of control points needed to define the curve of degree n and all the Bezier curves are differentiable with

continuous derivatives (De Adana *et al.*, 2011). Thus the Bezier parabola having degree 2 needs three control points to define the parabola.

The plane $\Pi(P_0, P_1, P_2)$ is defined by the three non co-linear points P_0, P_1 and P_2 in $\mathbb{R}^P$. Then the parameterized equation of the parabola is given by $M(t)$ in equation (4.1) or equivalently in (4.2).

$$M(t) = (1-t)^2 P_0 + 2(1-t)P_1 + t^2 P_2 \tag{4.1}$$

where the parameter t $\in [0,1]$

With $\Delta P_0 = P_1 - P_0$ and $\Delta P_1 = P_2 - P_1$ equation (4.2) is obtained from the equation (4.1).

$$M(t) = P_0 + 2t\Delta P_1 + t^2 \Delta^2 P_2 \tag{4.2}$$

where the parameter $t \in [0,1]$

When t is allowed to take values from the whole real line, the equation (4.1) gives the Bezier parabola. This gives a unique parabola which passes through the points P_0 and P_1 and is tangent to the lines l_1 and l_2 as shown in figure 4.6. The vector $\Delta^2 P_0$ directs the axis of the parabola.

The basic idea of PEM lies on the fact that three estimates of the parameters will control the local curvature of the surface consists of the parameters and the likelihood $(\theta, L(\theta))$ (Berlinet and Roland, 2012). Since the EM moves quickly closer to a neighborhood of a stationary point, it was attempted to use the Bezier parabola and then maximizing the likelihood over a subset of the parabola. Berlinet and Roland (2009) also have proved that the sequence of estimates generated by PEM increases likelihood.

The PEM algorithm starts similar to the general EM by accepting initial values for the parameters (P_0). Then it is requires to perform two iterations of the general EM to generate two estimates of the parameters (P_1 and P_2). The

three estimates P_0, P_1 and P_2 are then used as the control points of the parabola and define $M(t)$ as given in the equation (4.1) for $t \in \mathbb{R}$. Starting from P_2, a subset of the parabola is maximized in each iteration until the likelihood can not be further improved on the parabola. If the desired likelihood is achieved at the beginning, $M(t)$ is equal to P_2 for $t = 1$. Otherwise, starting from $t = 1$, the algorithm performs a geometric search on a grid to compute the increasing maximum likelihood at each iteration until the likelihood can not be increased any more (Berlinet and Roland, 2009). $\mathbf{R}$ implementation of the PEM is given in Appendix B.

The algorithm for PEM does not change the original structure of the EM, which enables a fair comparison to be made between PEM and the original EM that was implemented for the proposed model.

4.2.2 Efficiency in the Constant Model

In section 4.1 it was identified that the use of NR method is more efficient in place of $glm()$ function of R. As a further improvement the basic EM used in the previous section is replaced with PEM and efficiency was evaluated using simulations.

Data sets for the constant model were simulated using the true parameter values $p = .35, \beta_0^+ = -.5, \beta_1^+ = .2, \beta_0^- = -.7, \beta_1^- = -.1$. The execution time and the number of iterations were compared on both implementations of the EM (with NR method) and PEM algorithms, as shown in table 4.3.

Conforming to the work of Berlinet and Roland (2012), the PEM is more efficient in both execution time and the number of iterations. Figures 4.7 and 4.8 show the plots of the execution times and the number of iterations that are given in table 4.3. Up to the size 10^5, both EM and PEM, provided very close execution
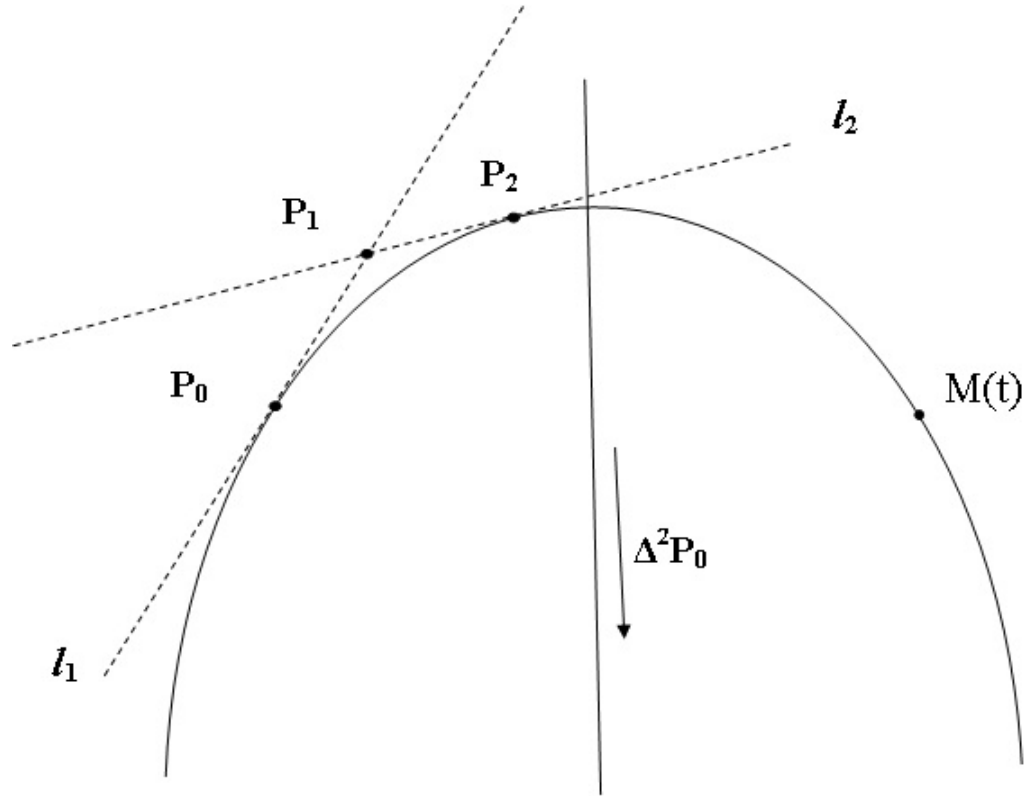
FIGURE 4.6 – Control Points P_0, P_1 and P_2 makes a triangle on the parabola

times. However, when the data set size was increased to be above 10^5, the PEM accelerated its execution. It can conclude that, for larger data sets PEM gives a better execution time than EM.

Also PEM cuts down the number of iterations by about one third. This means compared to EM, an iteration in PEM takes more time. The expected stability of the PEM was also achieved, with the failure rate of 0% in all the executions in the constant model.

4.2.3 Efficiency in the Variable Model

The data sets for variable model are generated similar to the constant model with $\alpha_0 = 0.3$ and $\alpha_1 = 0.8$ in the mixing parameter. The execution time and the

Constant Mixing Probability				
	EM with NR		PEM	
Size	Time	Iterations	Time	Iterations
100	0.05	33	0.05	12
1000	0.16	45	0.11	13
10000	1.01	43	0.94	14
100000	10.08	42	10.20	15
1000000	105.83	43	76.33	13

TABLE 4.3

Comparison of Efficiency with EM and PEM on Constant Model

number of iterations were compared on both implementations of EM (with NR method) and PEM algorithms, as shown in table 4.4. The figures 4.9 and 4.10 summarize the values given in table 4.4.

The test results of the simulations do not favor the PEM. The PEM was originally introduced for a mixture model with constant probability. Thus it is reasonable to expect its performance over constant probability, but not in variable mixture probabilities. As figures 4.9 and 4.10 show, the performance of PEM in the variable model is opposite to that of the constant model.

Apart from this poor performance, PEM was not stable during several executions. About 30% of the time, PEM failed to reach the maximum point even with 100000 iterations. Thus, it can be concluded that PEM is not very well suited for the proposed mixture model with variable mixing probabilities.
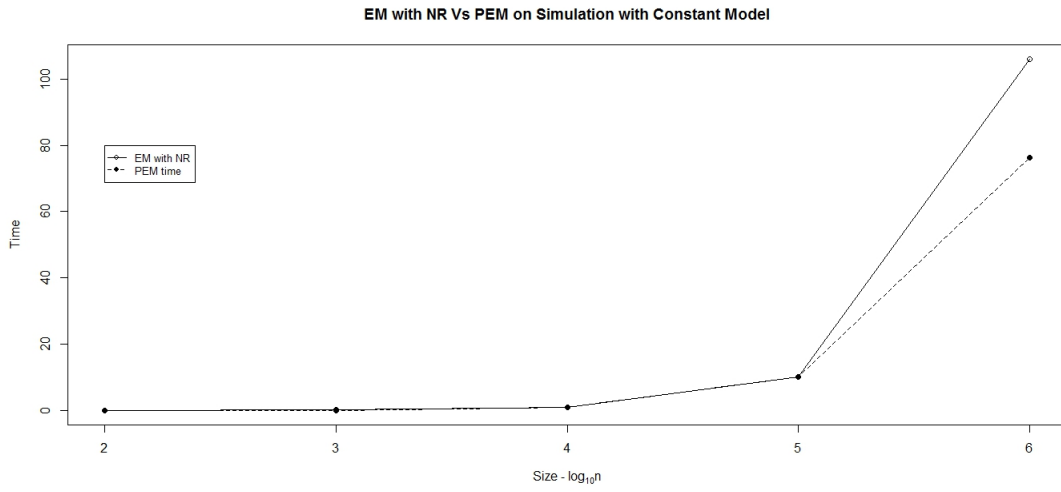
FIGURE 4.7 – Time of EM vs PEM for constant model on simulated data
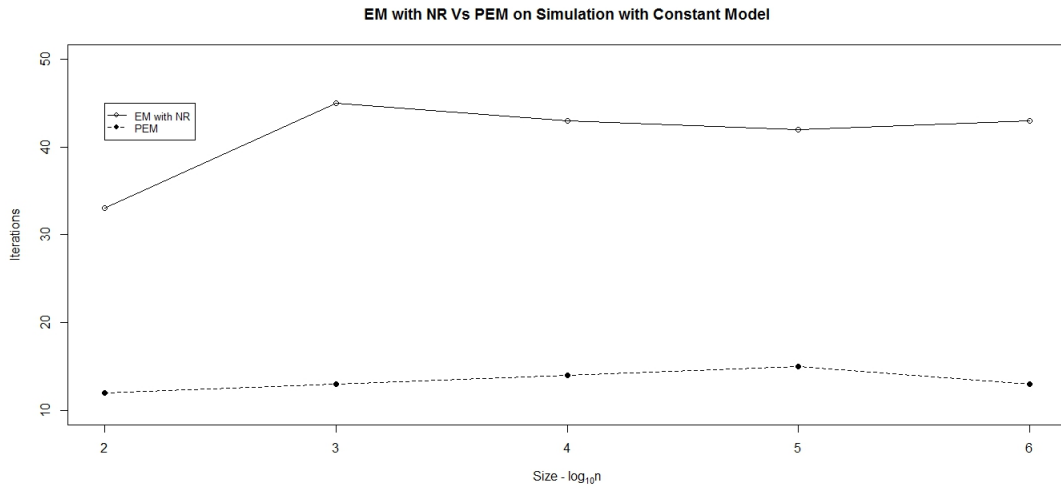
FIGURE 4.8 – Number of Iterations of EM vs PEM for constant model on simulated
data

74

Variable Mixing Probability				
	EM with NR		PEM	
Size	Time	Iterations	Time	Iterations
100	0.08	27	0.45	40
1000	0.18	32	0.55	58
10000	1.18	38	4.94	61
100000	10.36	32	45.58	58
1000000	111.64	32	452.48	58

TABLE 4.4

Comparison of Efficiency with EM and PEM on Variable Model
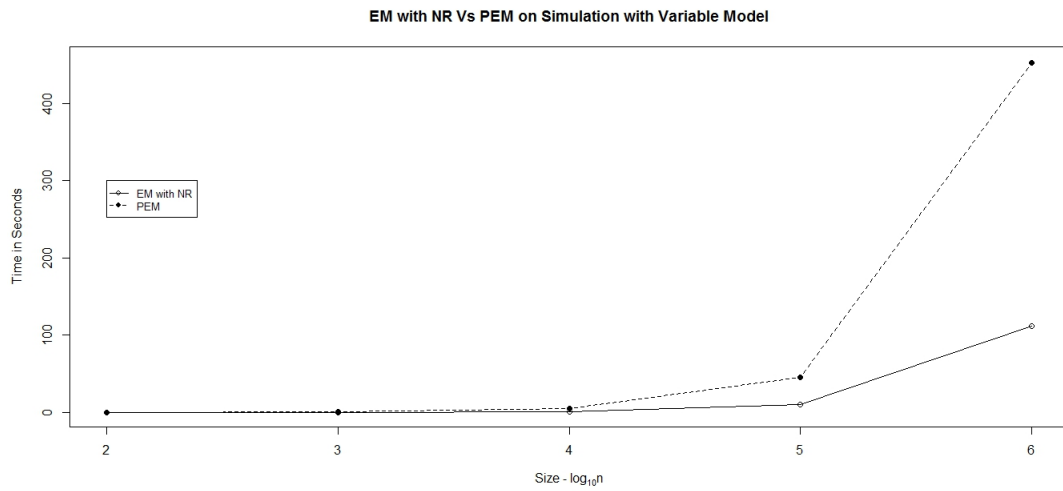
FIGURE 4.9 – Time of EM vs PEM for Variable model on simulated data
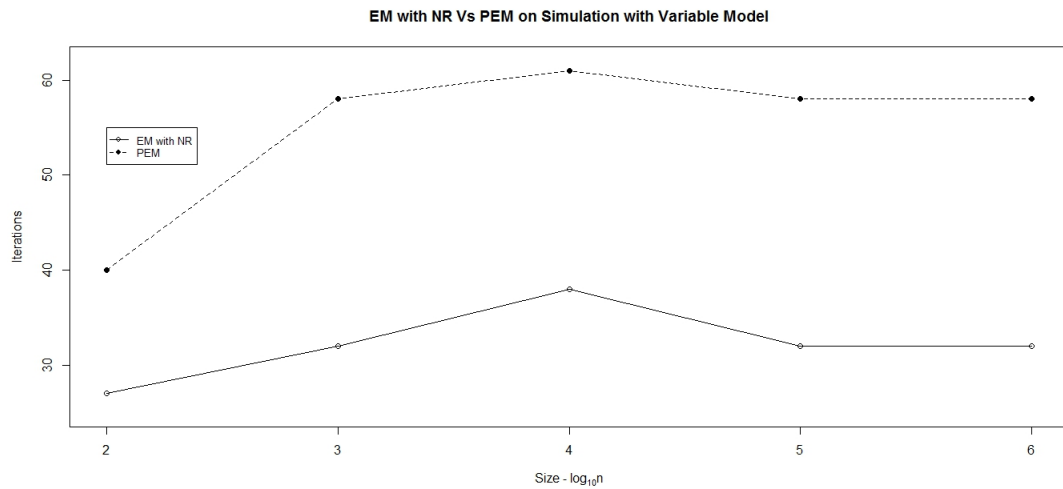
FIGURE 4.10 – Number of Iterations of EM vs PEM for Variable model on simu-
lated data

4.3 Parallel Processing

As another way of reducing the execution time of the model, a possibility of using a parallel processing environment was investigated. The High Performance Computing (HPC) facilities of the Cardinal Research Cluster (CRC) of the University of Louisville were utilized in achieving this task. Figure 4.11 shows the infrastructure of the HPC cluster.

The use of HPC cluster was beneficial when performing simulations with large amounts of data. The HPC cluster was accessed through a SSH Secure Shell within the university network. A Virtual Private Network (VPN) was needed to use the SSH Secure Shell when accessing from outside the university network.

The **R** codes of the models and the simulations were executed on 40 and 100 parallel processes to reduce the run time. Depending on the number of parallel processes used, whether it is 40 or 100, a list of seeds were generated and used in a separate file so that the same could be used if the outputs needed to be generated repeatedly under the same environment. Then additional codes were written using *Unix* commands to separated the codes into the number of processes and to combine the outputs once execution is completed.
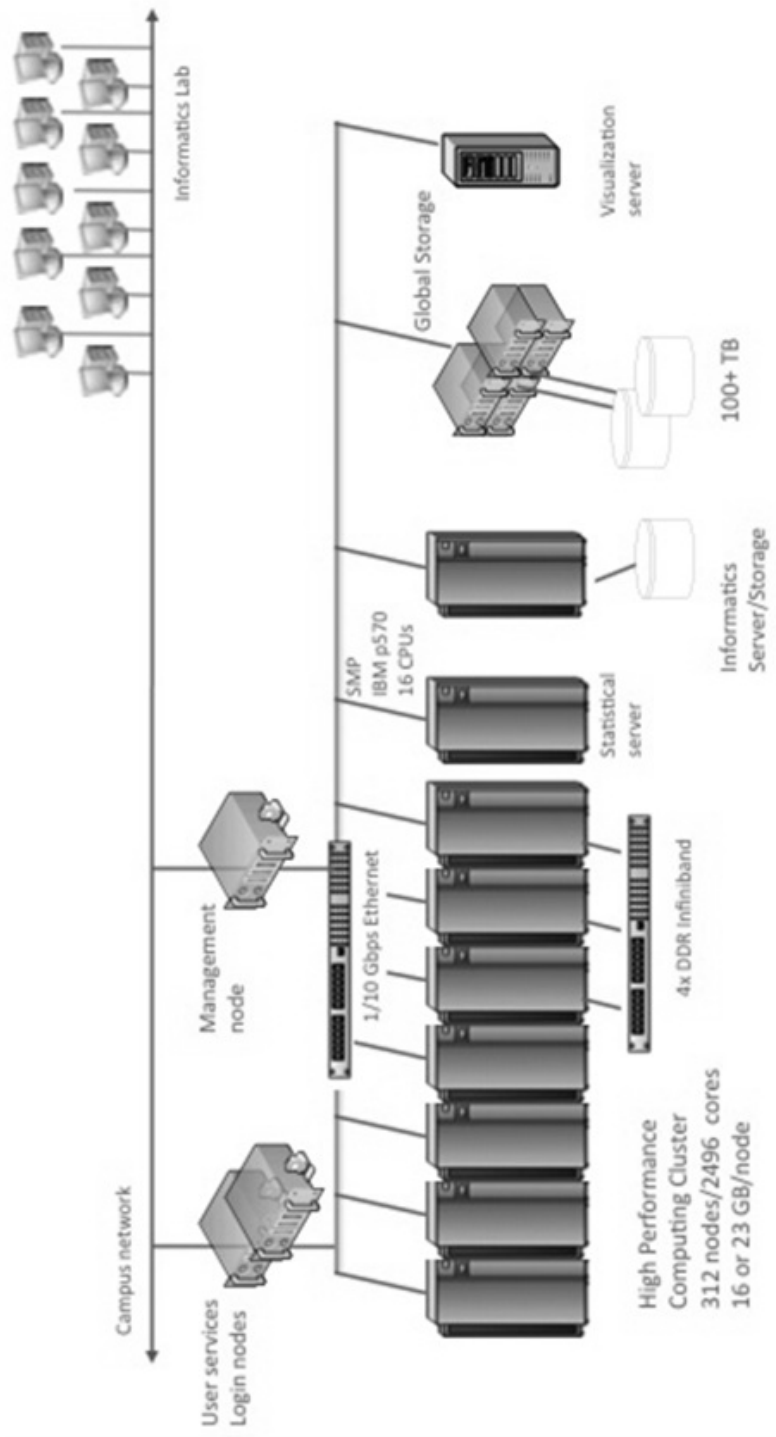
FIGURE 4.11 – HPC Cluster (source : http://louisville.edu/it/research/for-researchers/materials)

CHAPTER 5
CLUSTERED SIGNED MODEL

One of the biggest challenge in handling data is effectively manipulating large volumes of data. A similar challenge was experienced during the data manipulation for the proposed mixture model. When possible, it is desired to investigate if difficulties manipulating large volumes of data can be alleviated when the data set has special structures.

The proposed model employs the discrete stock price changes. The model considered those stock price changes in terms of negative, positive and also zero price changes. It was already discussed the effect of order size on the price change. According to the stock market the stocks are traded as a multiple of hundreds with hundred being the smallest size and also the most frequent size of a trade. Based on the data the order size do not have a large variation of different values. Therefore, it is possible to cluster the stock transactions based on different order sizes. Instead of taking each single stock price change and its order size, all the stock prices based on each different order size can be clustered. This clustering based on each order size summarizes data reducing the amount of repeatedly using multiple order sizes.

As per the proposed mixture model, the discrete stock price changes are treated based on their sign; negative, positive and zero. This fact is also used to further summarize data. Along with clustering, the respective discrete stock prices are summarized further, but taking sum of the magnitude of the sum of the stock price changes for sign from negative, positive and zeros.

A clustered model based on order size and the sign of the stock price change is introduced to reduce the amount of data used in the model. Section 5.1 presents the 'Clustered Signed Poisson Mixture Model' for constant mixture probability, and the section 5.2 presents the variable mixture probabilities.

5.1 Clustered Signed Poisson Mixture Model with Constant Mixing Probability

First consider the different discrete stock price changes made based on each different order size. Those are the observations clustered by order size given by x_i. For each value of x_i there are N observations with different discrete price changes given by y_{ij}. The observed independent data are given by $(x_1, y_{11}), \ldots, (x_1, y_{1N})$, $(x_2, y_{21}), \ldots, (x_2, y_{2N}), \cdots, (x_M, y_{M1}), \ldots, (x_M, y_{MN})$ where y_{ij} is a realization of the random variable Y_{ij}. The random variable Y_{ij} is defined using the mixture of two random variables Y_{ij}^- and Y_{ij}^+ as given by the expression (5.1).

$$Y_{ij} = (1 - \Delta_{ij})(-Y_{ij}^-) + \Delta_{ij}Y_{ij}^+ \tag{5.1}$$

with

$$\begin{aligned}
\Delta_{ij} &\sim \text{Bernoulli}(p) \\
Y_{ij}^+ &\sim \text{Poisson}(\lambda_i^+ = e^{\beta_0^+ + \beta_1^+ x_i}) \text{ and} \\
Y_{ij}^- &\sim \text{Poisson}(\lambda_i^- = e^{\beta_0^- + \beta_1^- x_i})
\end{aligned}$$

Once the model was set, the EM algorithm is applied for computing the estimates of the unknown model parameters. The step wise procedure for the EM algorithm is given from *Step* 1 to *Step* 4 below.

Step 1 : Start with initial estimates of $p, \beta_0^-, \beta_1^-, \beta_0^+, \beta_1^+$.

Step 2 : (**E-step**) Estimate responsibilities (estimates of $P(\Delta_{ij} = 1|y_{ij})$)

$$\gamma_{ij} = \begin{cases} 1 & \text{if } y_{ij} > 0 \\ \dfrac{pe^{-\lambda_i^+}}{pe^{-\lambda_i^+} + (1-p)e^{-\lambda_i^-}} & \text{if } y_{ij} = 0 \\ 0 & \text{if } y_{ij} < 0 \end{cases}.$$

Step 3 : (**M-step**) Maximize the complete likelihood with fixed gammas by solving the system of equations given in (5.2) to (5.6).

$$\sum_{i=1}^{M}\sum_{i=1}^{N}(1 - \gamma_{ij})\left\{-y_{ij} - e^{\beta_0^- + \beta_1^- x_i}\right\} = 0 \tag{5.2}$$

$$\sum_{i=1}^{M} x_i \sum_{i=1}^{N}(1 - \gamma_{ij})\left\{-y_{ij} - e^{\beta_0^- + \beta_1^- x_i}\right\} = 0 \tag{5.3}$$

$$\sum_{i=1}^{M}\sum_{i=1}^{N}\gamma_{ij}\left\{y_{ij} - e^{\beta_0^+ + \beta_1^+ x_i}\right\} = 0 \tag{5.4}$$

$$\sum_{i=1}^{M} x_i \sum_{i=1}^{N}\gamma_{ij}\left\{y_{ij} - e^{\beta_0^+ + \beta_1^+ x_i}\right\} = 0 \tag{5.5}$$

$$\frac{1}{MN}\sum_{i=1}^{M}\sum_{i=1}^{N}\gamma_{ij} = p \tag{5.6}$$

Step 4 : If the convergence of parameter estimates is not achieved in two consecutive steps, go back to step 2.

The M-step requires *Weighted Poisson regression* to solve some of the equations. The Weighted Poisson regression with weights w_i, sizes n_i, and data (x_i, y_i) for $i = 1, \ldots, n$ computes the values θ_0 and θ_1 which solve the system of equations given in (5.7) and (5.8).

$$\sum_{i=1}^{N} w_i \left\{y_i - e^{\theta_0 + \theta_1 x_i}\right\} = 0 \tag{5.7}$$

$$\sum_{i=1}^{N} x_i w_i \left\{y_i - e^{\theta_0 + \theta_1 x_i}\right\} = 0 \tag{5.8}$$

The equations given in (5.2) and (5.6) are simplified further to compute using Poisson regression. As a way of improving the efficiency of the EM algorithm, this simplification attempts to reduce the amount of data used for the E-step.

There are a large number of transactions with the discrete stock price change and its order size. There are also a significant number of transactions those were based on similar order size. It is useful to consider the clusters of transactions based on each different order size. With that, for each different x_i there are a number of y_{ij}'s. Under the clustered signed model, the amount of data that is used for the mixture model is reduced by considering the summarized data based on each different x_i, instead of considering all the (x_i, y_{ij}) pairs.

The required notation is described below.

Let,

M = number of distinct x_i values

N_{i0} = number of y_{ij}'s that equal 0,

N_{i+} = number of y_{ij}'s that are positive, and

N_{i-} = number of y_{ij}'s that are negative.

Then N_{i0} of the γ_{ij}'s are in the interval $(0, 1)$, N_{i+} of the γ_{ij}'s equal 1, and N_{i-} of the γ_{ij}'s equal 0 where

$$\gamma_{i0} = \frac{pe^{-\lambda_i^+}}{pe^{-\lambda_i^+} + (1-p)e^{-\lambda_i^-}},$$

Also let,

y_{i+} = sum of the positive y_{ij}'s and

y_{i-} = absolute value of the sum of the negative y_{ij}'s.

Then the equations (5.2) to (5.6) can be rewritten as the equations given by (5.9) to (5.13).

$$\sum_{i=1}^{M} N_{i-} \left\{ \frac{y_{i-}}{N_{i-}} - e^{\beta_0^- + \beta_1^- x_i} \right\} + \sum_{i=1}^{M} N_{i0}(1 - \gamma_{i0}) \left\{ 0 - e^{\beta_0^- + \beta_1^- x_i} \right\} = 0 \quad (5.9)$$

$$\sum_{i=1}^{M} x_i N_{i-} \left\{ \frac{y_{i-}}{N_{i-}} - e^{\beta_0^- + \beta_1^- x_i} \right\} + \sum_{i=1}^{M} x_i N_{i0}(1 - \gamma_{i0}) \left\{ 0 - e^{\beta_0^- + \beta_1^- x_i} \right\} = 0 \quad (5.10)$$

$$\sum_{i=1}^{M} N_{i+} \left\{ \frac{y_{i+}}{N_{i+}} - e^{\beta_0^+ + \beta_1^+ x_i} \right\} + \sum_{i=1}^{M} N_{i0}\gamma_{i0} \left\{ 0 - e^{\beta_0^+ + \beta_1^+ x_i} \right\} = 0 \quad (5.11)$$

$$\sum_{i=1}^{M} x_i N_{i+} \left\{ \frac{y_{i+}}{N_{i+}} - e^{\beta_0^+ + \beta_1^+ x_i} \right\} + \sum_{i=1}^{M} x_i N_{i0}\gamma_{i0} \left\{ 0 - e^{\beta_0^+ + \beta_1^+ x_i} \right\} = 0 \quad (5.12)$$

$$\frac{1}{MN} \sum_{i=1}^{M} \{N_{i+} + N_{i0}\gamma_{i0}\} = p \quad (5.13)$$

Then the model parameters are estimated using Newton-Raphson algorithm.

5.2 Clustered Signed Poisson Mixture Model with Variable Mixing Probabilities

Similar to section 5.1, the Clustered Signed Poisson Mixture Model for variable mixing probabilities is defined with the observed independent data (x_1, y_{11}), $\ldots, (x_1, y_{1N}), (x_2, y_{21}), \ldots, (x_2, y_{2N}), \cdots, (x_M, y_{M1}), \ldots, (x_M, y_{MN})$ where y_{ij} is a realization of the random variable Y_{ij} which is given in equation (5.14).

$$Y_{ij} = (1 - \Delta_{ij})(-Y_{ij}^-) + \Delta_{ij} Y_{ij}^+ \quad (5.14)$$

with

$$\Delta_{ij} \sim \text{Bernoulli}(p_i = \frac{e^{\alpha_0 + \alpha_1 x_i}}{1 + e^{\alpha_0 + \alpha_1 x_i}})$$

$$Y_{ij}^+ \sim \text{Poisson}(\lambda_i^+ = e^{\beta_0^+ + \beta_1^+ x_i}) \text{ and}$$

$$Y_{ij}^- \sim \text{Poisson}(\lambda_i^- = e^{\beta_0^- + \beta_1^- x_i})$$

The EM algorithm for computing the estimates of the unknown model parameters is given below.

Step 1 : Start with initial estimates of $\alpha_0, \alpha_1, \beta_0^-, \beta_1^-, \beta_0^+, \beta_1^+$.

Step 2 : (**E-step**) Estimate responsibilities (estimates of $P(\Delta_{ij} = 1|y_{ij})$)

$$
\gamma_{ij} = \begin{cases} 1 & \text{if } y_{ij} > 0 \\ \dfrac{p_i e^{-\lambda_i^+}}{p_i e^{-\lambda_i^+} + (1-p_i) e^{-\lambda_i^-}} & \text{if } y_{ij} = 0 \\ 0 & \text{if } y_{ij} < 0 \end{cases}.
$$

Step 3 : (**M-step**) Maximize the complete likelihood with fixed gammas by solving the following system of equations.

Based on the clustering of order size the independent observations (x_1, y_{11}), $\dots, (x_1, y_{1N}), (x_2, y_{21}), \dots, (x_2, y_{2N}), \cdots, (x_M, y_{M1}), \dots, (x_M, y_{MN})$ are clustered as given in equations (5.15) to (5.20).

$$
\sum_{i=0}^{M} \sum_{i=0}^{N} (1 - \gamma_{ij}) \left\{ -y_{ij} - e^{\beta_0^- + \beta_1^- x_i} \right\} = 0 \tag{5.15}
$$

$$
\sum_{i=0}^{M} x_i \sum_{i=0}^{N} (1 - \gamma_{ij}) \left\{ -y_{ij} - e^{\beta_0^- + \beta_1^- x_i} \right\} = 0 \tag{5.16}
$$

$$
\sum_{i=0}^{M} \sum_{i=0}^{N} \gamma_{ij} \left\{ y_{ij} - e^{\beta_0^+ + \beta_1^+ x_i} \right\} = 0 \tag{5.17}
$$

$$
\sum_{i=0}^{M} x_i \sum_{i=0}^{N} \gamma_{ij} \left\{ y_{ij} - e^{\beta_0^+ + \beta_1^+ x_i} \right\} = 0 \tag{5.18}
$$

$$
\sum_{i=0}^{M} \sum_{i=0}^{N} \left\{ \gamma_{ij} - \frac{e^{\alpha_0 + \alpha_1 x_i}}{1 + e^{\alpha_0 + \alpha_1 x_i}} \right\} = 0 \tag{5.19}
$$

$$
\sum_{i=0}^{M} x_i \sum_{i=0}^{N} \left\{ \gamma_{ij} - \frac{e^{\alpha_0 + \alpha_1 x_i}}{1 + e^{\alpha_0 + \alpha_1 x_i}} \right\} = 0 \tag{5.20}
$$

Step 4 : If the convergence of parameter estimates is not achieved in two consecutive steps, go back to step 2.

The equations formulated in the M-step, are solved using either weighted Poisson regression or a weighted modification of logistic regression. The *weighted Poisson regression* with weights w_i, sizes n_i, and data (x_i, y_i) for $i = 1, \dots, n$

computes the values θ_0 and θ_1 which solve the system of equations given in the equations (5.21) and (5.22).

$$\sum_{i=0}^{n} w_i \left\{ y_i - n_i e^{\theta_0 + \theta_1 x_i} \right\} = 0 \tag{5.21}$$

$$\sum_{i=0}^{n} x_i w_i \left\{ y_i - e^{\theta_0 + \theta_1 x_i} \right\} = 0 \tag{5.22}$$

The *logistic regression* with sizes n_i and data (x_i, y_i) for $i = 1, \ldots, n$ computes the values θ_0 and θ_1 which solve the system of equations that are given by equations (5.23) and (5.24).

$$\sum_{i=0}^{n} \left\{ y_i - n_i \frac{e^{\theta_0 + \theta_1 x_i}}{1 + e^{\theta_0 + \theta_1 x_i}} \right\} = 0 \tag{5.23}$$

$$\sum_{i=0}^{n} x_i \left\{ y_i - n_i \frac{e^{\theta_0 + \theta_1 x_i}}{1 + e^{\theta_0 + \theta_1 x_i}} \right\} = 0 \tag{5.24}$$

Next the equations (5.15) to (5.20) are simplified using the clustered order sizes x_i and the signed discrete price changes for the E-step. Then the estimates computed using poisson and logistic regression.

Let

N_{i0} = number of y_{ij}'s that equal 0,

N_{i+} = number of y_{ij}'s that are positive, and

N_{i-} = number of y_{ij}'s that are negative.

Based on the model settings N_{i0} of the γ_{ij}'s are in the interval $(0, 1)$, N_{i+} of the γ_{ij}'s equal 1, and N_{i-} of the γ_{ij}'s equal 0 from

$$\gamma_{i0} = \frac{p_i e^{-\lambda_i^+}}{p_i e^{-\lambda_i^+} + (1 - p_i) e^{-\lambda_i^-}}.$$

Also let, y_{i+} = sum of the positive y_{ij}'s and

y_{i-} = absolute value of the sum of the negative y_{ij}'s.

Then the equations (5.15) to (5.20) can be rewritten as the equations given in (5.25) to (5.30).

$$\sum_{i=0}^{M} N_{i-} \left\{ \frac{y_{i-}}{N_{i-}} - e^{\beta_0^- + \beta_1^- x_i} \right\} + \sum_{i=0}^{M} N_{i0}(1 - \gamma_{i0}) \left\{ 0 - e^{\beta_0^- + \beta_1^- x_i} \right\} = 0 \quad (5.25)$$

$$\sum_{i=0}^{M} x_i N_{i-} \left\{ \frac{y_{i-}}{N_{i-}} - e^{\beta_0^- + \beta_1^- x_i} \right\} + \sum_{i=0}^{M} x_i N_{i0}(1 - \gamma_{i0}) \left\{ 0 - e^{\beta_0^- + \beta_1^- x_i} \right\} = 0 \quad (5.26)$$

$$\sum_{i=0}^{M} N_{i+} \left\{ \frac{y_{i+}}{N_{i+}} - e^{\beta_0^+ + \beta_1^+ x_i} \right\} + \sum_{i=0}^{M} N_{i0}\gamma_{i0} \left\{ 0 - e^{\beta_0^+ + \beta_1^+ x_i} \right\} = 0 \quad (5.27)$$

$$\sum_{i=0}^{M} x_i N_{i+} \left\{ \frac{y_{i+}}{N_{i+}} - e^{\beta_0^+ + \beta_1^+ x_i} \right\} + \sum_{i=0}^{M} x_i N_{i0}\gamma_{i0} \left\{ 0 - e^{\beta_0^+ + \beta_1^+ x_i} \right\} = 0 \quad (5.28)$$

$$\sum_{i=0}^{M} \left\{ (N_{i+} + N_{i0}\gamma_{i0}) - N \frac{e^{\alpha_0 + \alpha_1 x_i}}{1 + e^{\alpha_0 + \alpha_1 x_i}} \right\} = 0 \quad (5.29)$$

$$\sum_{i=0}^{M} x_i \left\{ (N_{i+} + N_{i0}\gamma_{i0}) - N \frac{e^{\alpha_0 + \alpha_1 x_i}}{1 + e^{\alpha_0 + \alpha_1 x_i}} \right\} = 0. \quad (5.30)$$

It should be noted that any large data set of size n is converted to a clustered signed model based on M distinct values of the order size x_i. Thus the sum of a large n has become a sum of M. Another important fact is that the clustered signed model is built based on the characteristics of a typical data set of tick–by–tick stock transactions, where there are multiple transactions with many distinct order sizes. Hence, will not be efficient for data sets where values of order sizes are not repeated.

Comparison of efficiency of the proposed clustered signed model with both constant and variable mixture probabilities are presented in section 7.3.

5.3 Efficiency

The clustered signed models proposed in sections 5.1 and 5.2 were implemented using **R**. Their execution times were compared with the implementations of the mixture model discussed in section 4.1 and the PEM in section 4.2. Table 5.1 and 5.2 show the execution times of the clustered model and mixture model with both constant and variable mixture probabilities. The figures 5.1, 5.2 and 5.3 graphically show the time efficiency of the clustered model.

In the implementation of clustered model, an additional functionality was required to summarize the order sizes (x_i) and discrete stock price changes to be used. The summarized values include clustered x_i values (xc_i), the number of y_i values that are zeros (N_{i0}), the number of y_i values that are negative (N_{i-}), the number of y_i values that are positive (N_{i+}), the sum of positive y_i values (y_{i+}) and the absolute sum of negative y_i (y_{i-}). The time shown in the above tables and figures for the clustered model also includes the time for summarization of data. A sample of processed data is given in Appendix A.3.

Outputs clearly show there is a significant gain in time when the clustered model is used in compared to the mixture model proposed in chapter 3. The clustered model shows a more efficient time compared to both the improved implementation of the EM algorithm in section 4.1 (figures 5.1 and 5.3) and the PEM algorithm version for constant model explained in section 4.2 (figure 5.2).

Stock transactions data consist of clustered values both order sizes (x_i) and (y_i) price changes. This was further conformed during the data analysis in chapter 2. For an example, there are significant number of orders placed with the order sizes ± 100 followed by ± 200 and ± 300. Thus, the suitability of clustered model well reasonable.

Size	Clustered Model	Mixture Model
100	0.01	0.01
1000	0.01	0.02
10000	0.04	1.47
100000	0.21	14.65
1000000	1.91	142.57

TABLE 5.1

Execution times (in seconds) of Clustered Signed Model and Mixture Model with constant probability

Size	Clustered Model	Mixture Model
100	0.01	0.02
1000	0.02	0.24
10000	0.04	3.12
100000	0.22	28.54
1000000	3.43	303.42

TABLE 5.2

Execution times (in seconds) of Clustered Signed Model and Mixture Model with variable mixture probability
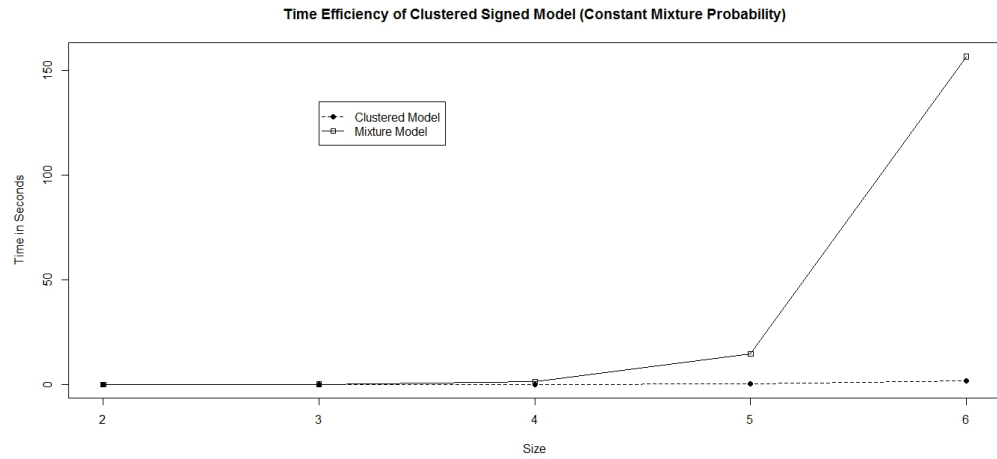
FIGURE 5.1 – Time comparison of Clustered Signed Model and Mixture Model with constant probability
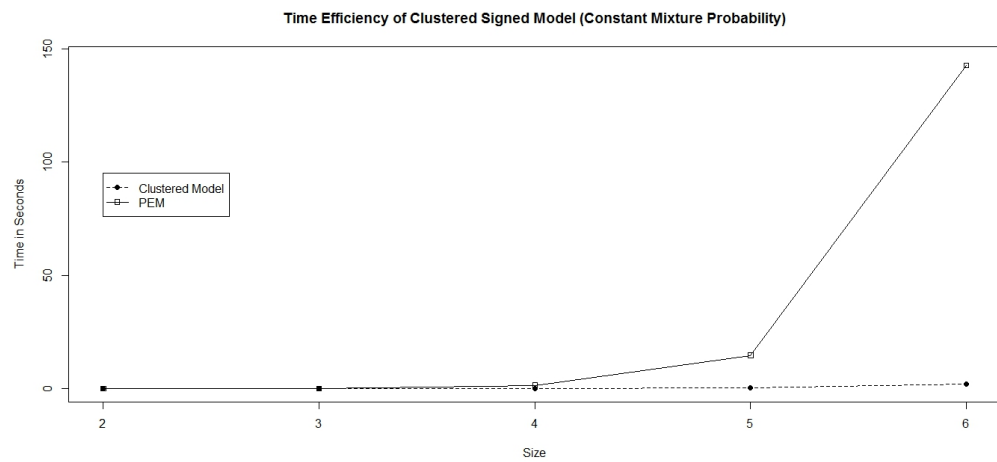
FIGURE 5.2 – Time comparison of Clustered Signed Model and PEM algorithm with constant probability

Time Efficiency of Clustered Signed Model (Variable Mixture Probability)
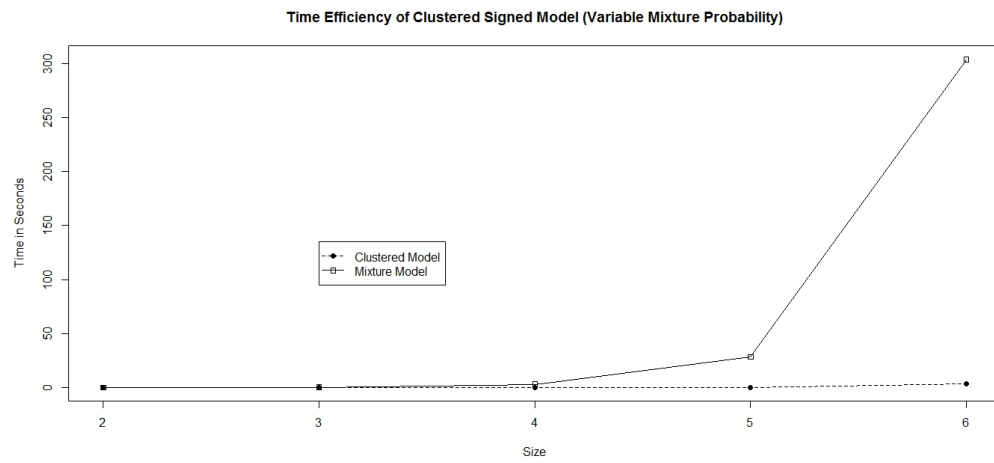
FIGURE 5.3 – Time comparison of Clustered Signed Model and Mixture Model with variable probability

CHAPTER 6
TEST FOR MIXTURE PROBABILITY

Two versions of the poisson mixture model were proposed in chapter 3. The initial model consists of a constant mixture probability (p) which is commonly found in mixture model. The extension was the variable mixture probability (p_i) that depends on the order size (x_i). As the order size plays a major role in the stock price change, it seems that the variable mixture probability is a reasonable extension of the model. However, it is important to determine whether the data has a strong evidence to support the use of the model with variable mixing probabilities compared to the model with the constant mixture probabilities.

The actual distribution of the proposed model is complex. In such cases, bootstrap methods make the test of significance easier to compute. Bootstrap methods are not asymptotic procedures. Thus work independently to asymptotic theories. A significance test using a 'Parametric Bootstrap' method is decided to perform and the results are presented in following sections.

6.1 Parametric Bootstrap

According to Chernick (1999), bootstrap means re-sampling from an original data set. Methods of bootstrapping are also called 're-sampling procedures'. As Martinez-Camblor and Corral (2012) state, bootstrap methods are useful when measuring accuracy to statistical estimates.

The first bootstrap procedure was introduced by Bradley Efron in 1979,

which focuses on 'Non-parametric bootstrap'. As Chernick (1999) outlines, the formal definition of Efron's bootstrap (Non-parametric bootstrap), is given in the definition 6.2. It is also important to distinguish 'Parametric' and 'Non-Parametric' models. The definitions of parametric and non-parametric models as stated by Davison and Hinkley (1997) are given in definition 6.1.

Zhu (1997) highlights that it is sometimes better to make conclusion about the population parameters based on the samples drawn from the original sample than using the population to make unrealistic assumptions. Zhu (1997) also mentions that when the formula for the population parameters are not available, the bootstrapping provide a useful alternative. Nevertheless, according to Zhu (1997), the bootstrapping will not be a good solution when the original sample does not represent the population very well or for highly skewed populations.

DEFINITION 6.1 (Parametric and Non-Parametric Models). *A Mathematical model is called parametric, when there is a fully determined probability density function with adjustable constants or parameters, is available for the model.*

In the absence of such fully determined probability density functions, the statistical analysis uses only the fact that the random variables are independent and identically distributed. Thus are called non-parametric.

DEFINITION 6.2 (Non-Parametric Bootstrap). *Let $X_1, X_2, \ldots, X_n$ denote a sample of n independent identically distributed random vectors and $\hat{\theta} = \theta(X_1, X_2, \ldots, X_n)$ denotes the real valued estimator of the distribution of parameter θ. An empirical distribution function F_n is used in a bootstrap procedure to assess the accuracy of $\hat{\theta}$. The probability of $\frac{1}{n}$ is assigned to each observed values of the random vectors X_i for $i = 1, 2, \ldots, n$, by the assumed empirical distribution function F_n.*

Chernick (1999) further states that, according to the 'Strong Law of Large Numbers' for independent and identically distributed random samples, the function

F_n as given in the definition 6.2 converges to F point-wise with probability one.

In the non-parametric bootstrap the data distribution serves as the empirical distribution. The concept of parametric bootstrap is similar to the non-parametric bootstrap. In non-parametric bootstrap, the bootstrap samples are simulated from the empirical distribution of the independent identically distributed data, where as, the bootstrap samples for parametric bootstrap are simulated from an assumed parametric distribution. The parametric bootstrap is preferred when a properly specified model is used for the application. However, in both the bootstrap methods, a larger sample sizes are used to improve the accuracy of the estimation.

In order to evaluate the significance of the use of the mixture probability in the model, a 'Hypothesis Test' will be used. 'Hypothesis Testing' requires handing two sampling distributions (Shalizi, 2011), one under the null hypothesis and the other under the alternative. Shalizi (2011) further states that the size of the test and the significance level is obtained by the test statistic under the null hypothesis and the power and realized power of the test is obtained by the alternative. Martinez-Camblor and Corral (2012) states that "the bootstrap methods provide a creative way for building hypothesis testing without the need for restrictive parametric assumptions".

6.2   Significance Test for $\alpha_1 = 0$

A hypothesis test is proposed to assess whether the significance of the order size on the mixing probabilities. The test proposed in this section uses the 'Clustered Signed Model' proposed in chapter 5 and the estimates obtained from the EM algorithm that was described in chapter 5 based on both models. The parametric bootstrap is used to test whether the magnitude of α_1 is significantly different from 0.

Under the proposed model the mixtures are assumed to be following poisson distributions. Thus the model is a properly specified. This fact is used in building the parametric bootstrap method, with 1000 bootstrap samples. The next section explains the use of parametric bootstrap in performing the significance test.

6.2.1 The Significance Test

The variable mixture probability that depends on the order size has two parameters (α_0 and α_1) as given by the equation 6.1.

$$p_i = \frac{\exp(\alpha_0 + \alpha_1 x_i)}{1 + \exp(\alpha_0 + \alpha_1 x_i)} \tag{6.1}$$

If the α_1 parameter is 0, then the effect of order size (x_i) in the mixture probability will vanish and thus the problem converts the use of only the constant mixture probability.

The test statistic is defined with the null hypothesis $H_0 : \alpha_1 = 0$ versus the alternative hypothesis $H_a : \alpha_1 \neq 0$ for the likelihood $l(\alpha_0, \alpha_1, \beta_0^-, \beta_1^-, \beta_0^+, \beta_1^+)$ under the alternative hypothesis. The generalized test statistic for the test is

$$\Lambda = \frac{sup_{H_0 \cup H_a} \, l}{sup_{H_0} \, l} \tag{6.2}$$

with the rejection rule, where H_0 is rejected if Λ is sufficiently larger than the critical value given by Λ^*. That is when $\Lambda > \Lambda^*$ the null hypothesis H_0 is rejected.

Under the null hypothesis H_0 with $\alpha_1 = 0$, the mixture probability p becomes a constant and given by $p = \frac{e^{\alpha_0}}{1+e^{\alpha_0}}$. This gives that $\alpha_0 = ln(\frac{1-p}{p})$ under the null hypothesis H_0. This leads to the estimation of the parameters $\tilde{p}, \tilde{\beta}_0^-, \tilde{\beta}_1^-, \tilde{\beta}_0^+, \tilde{\beta}_1^+$ under the constant model. The alternative hypothesis $H_a : \alpha_1 \neq 0$ results the use of the model with the variable mixing probability to find the parameters $\hat{\alpha}_0, \hat{\alpha}_1, \hat{\beta}_0^-, \hat{\beta}_1^-, \hat{\beta}_0^+, \hat{\beta}_1^+$.

Then the observed values are used to compute Λ_{obs} based on the hypothesis

as given in the expression (6.3).

$$\Lambda_{obs} = \frac{l(\hat{\alpha}_0, \hat{\alpha}_1, \hat{\beta}_0^-, \hat{\beta}_1^-, \hat{\beta}_0^+, \hat{\beta}_1^+)}{l(\tilde{\alpha}_0, 0, \tilde{\beta}_0^-, \tilde{\beta}_1^-, \tilde{\beta}_0^+, \tilde{\beta}_1^+)} \tag{6.3}$$

Then the bootstrap principle is applied using the estimates based on the null model to generate B bootstrap samples. B number of bootstrap samples $\{(x_1, y_1^{(b)}), \ldots, (x_n, y_n^{(b)})\}$, $b = 1, \ldots, B$ were generated and for each bootstrap sample denoted by b, the estimates under both models. The bth bootstrap sample is generated from the mixture under the null model as follows.

First, a latent variable Δ_{bi} is generated from a Bernoulli distribution with mean $\hat{p}$. If $\Delta_{bi} = 0$, then y_{bi} is generated from a Poisson distribution with mean $\exp(\tilde{\beta}_0^- + \tilde{\beta}_1^- x_i)$; otherwise, if $\Delta_{bi} = 1$, then y_{bi} is generated from a Poisson distribution with mean $\exp(\tilde{\beta}_0^+ + \tilde{\beta}_1^+ x_i)$.

The $\tilde{p}^{(b)}, \tilde{\beta}_0^{-(b)}, \tilde{\beta}_1^{-(b)}, \tilde{\beta}_0^{+(b)}, \tilde{\beta}_1^{+(b)}$ represents the estimates under the constant model whereas $\hat{\alpha}_0^{(b)}, \hat{\alpha}_1^{(b)}, \hat{\beta}_0^{-(b)}, \hat{\beta}_1^{-(b)}, \hat{\beta}_0^{+(b)}, \hat{\beta}_1^{+(b)}$ represents the estimates under the variable model.

Then for each bootstrap b, the value Λ^b is computed based on the expression given in (6.4).

$$\Lambda^b = \frac{l(\hat{\alpha}_0^{(b)}, \hat{\alpha}_1^{(b)}, \hat{\beta}_0^{-(b)}, \hat{\beta}_1^{-(b)}, \hat{\beta}_0^{+(b)}, \hat{\beta}_1^{+(b)})}{l(\tilde{\alpha}_0^{(b)}, 0, \tilde{\beta}_0^{-(b)}, \tilde{\beta}_1^{-(b)}, \tilde{\beta}_0^{+(b)}, \tilde{\beta}_1^{+(b)})} \tag{6.4}$$

For each b the $\widehat{p-value}$ is computed using the expression given in (6.5).

$$\widehat{p-value} = \frac{\text{Number of times } \Lambda_{obs} < \Lambda^{(b)}}{B} \tag{6.5}$$

and the null hypothesis is rejected if the estimated p-value is less than a pre-specified significance level.

6.2.2 Simulation Results

The proposed significant test is implemented as a simulation to assess the size and power of the parametric bootstrap procedure. The simulations were per-

formed with the true parameter values $\alpha_0 = 0.3$, $\beta_0^+ = -0.5$, $\beta_1^+ = 0.2$, $\beta_0^- = -0.7$, and $\beta_1^- = -0.1$. For each simulated data set, r number of observations were used with each at order sizes $-5, -4, -3, -2, -1, 1, 2, 3, 4, 5$. The model with constant mixing probabilities is equivalent to the model with variable mixing probabilities with $\alpha_1 = 0$.

The power of the test, that is the probability of correctly rejecting the null hypothesis when it is false, was calculated for different α_1 values. Table 6.1 shows the estimated power of the tests for different number of observations and α_1 values. It can be observed that the power of α_1 values that are closer to 0 is closer to the significance level. When the α_1 values are significantly different from zero, the power moves further away from the significance level.

The power of the tests given in table 6.1 is graphed in figure 6.1. For symmetric α_1 values, the graph looks symmetric as expected with the minimum around the significance level 0.05 and increased to the maximum power 1.

Based on the results from the above test, it can be concluded that the test was performed as expected. When $\alpha_1 = 0$, the estimated power of the test was very close to 5%. In other words, according to the test, the null hypothesis should only be rejected about 5% of the time. Also, for each fixed r, the power increases as the magnitude of α_1 increases. For each fixed non-zero α_1, the power increases as the number of observations increase.

α_1	Estimated power for		
	$r = 100$	$r = 200$	$r = 500$
-0.250	1.000	1.000	1.000
-0.125	0.847	0.994	1.000
-0.100	0.635	0.942	0.989
-0.075	0.395	0.702	0.989
-0.050	0.187	0.392	0.556
-0.025	0.069	0.117	0.279
0.000	0.055	0.051	0.053
0.025	0.149	0.193	0.371
0.050	0.312	0.556	0.886
0.075	0.597	0.859	0.996
0.100	0.816	0.981	1.000
0.125	0.930	0.997	1.000
0.250	1.000	1.000	1.000

TABLE 6.1

Estimated power for tests based on parametric bootstrap at significance level 0.05 based on 1000 simulations of size r.
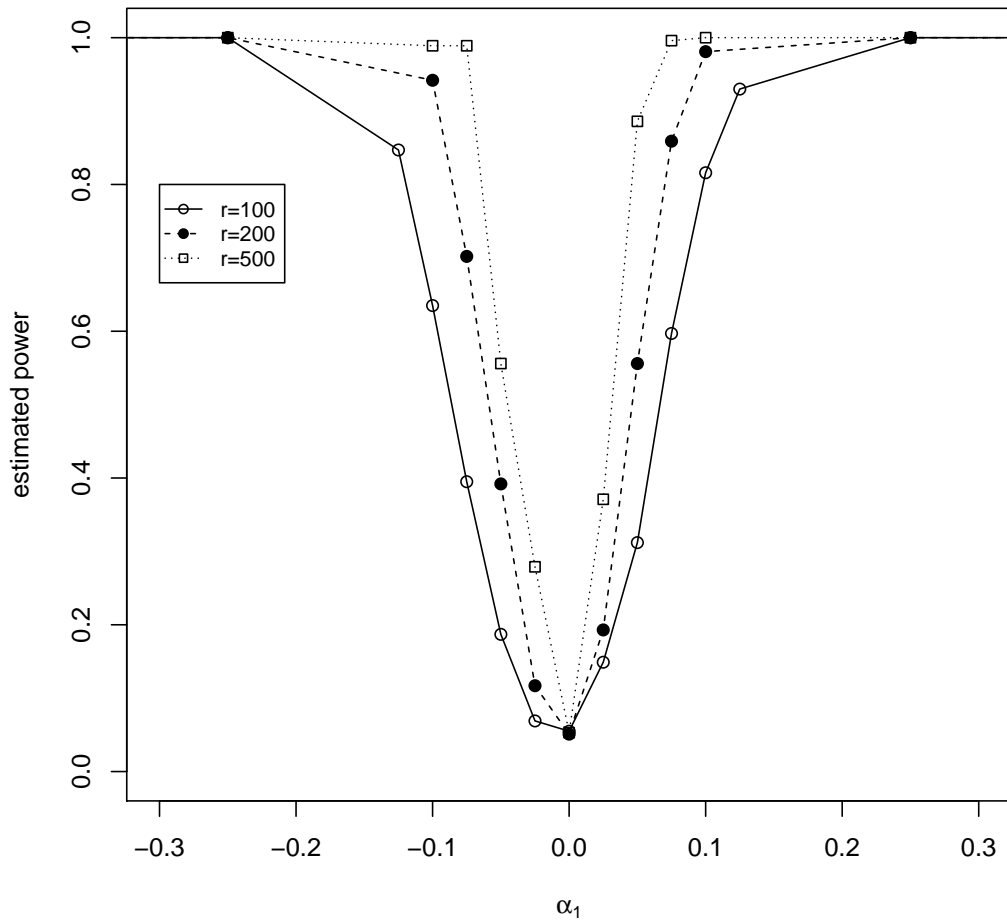
FIGURE 6.1 – Estimated power curves for parametric bootstrap procedure at significance level 0.05 based on 1000 simulati The solid line is for r=100, the dotted line with solid points is for r=200 and the dotted lines with squares for the point is for r=500.

CHAPTER 7
APPROXIMATE CONFIDENCE INTERVAL

The significance test performed in Chapter 6, confirms the appropriateness of the variable mixture probability in the proposed mixture model. The next step during this chapter is to evaluate the approximate confidence interval for the variable mixture probability p_i and some interesting probabilities which were presented in section 3.3.2.

One of the most popular method of finding confidence intervals, 'Delta Method', is used to find approximate confidence interval. Section 7.1 briefly outlines the Delta Method, as described by Uusipaikka (2008).

7.1   Delta Method

Let $g(\theta)$ be a real valued function of interest and r be the value of the function $g(\theta)$. Let $J(\theta)$ is the observed information matrix. Then the confidence interval of r as produced by the delta method is given in the expression (7.1).

$$r \in \hat{r} \mp Z_{A/2}^* \sqrt{\frac{\partial g(\hat{\theta})^T}{\partial \theta} J(\hat{\theta}) \frac{\partial g(\hat{\theta})}{\partial \theta}} \tag{7.1}$$

where $\hat{r} = g(\hat{\theta})$ is the maximum likelihood estimate of r and θ denotes the population parameters.

The observed information function, $J(\theta)$, is found by negating the second derivative of the likelihood function.

7.2   Log Likelihood

It is important to identify that the log–likelihood given in expression (3.4) of the proposed mixture model consists of a logarithm of a sum of two components from the two sub populations. The logarithm of the sum makes the log–likelihood complex. Thus computing estimators from the log–likelihood is extremely difficult. Therefore, EM algorithm was used as a solution when finding the estimates. However, the log–likelihood is required to use directly when finding the confidence intervals for population parameters.

Due to the complexity of the log-likelihood, the logarithm of the sum is needed to be adjusted. Czado and Min (2005) have used a trick in simplifying a logarithm of a sum in a similar log–likelihood function of a 'Zero-Inflated Generalized Poisson Model' to divide the logarithm of a sum into a sum of three logarithms. A similar trick is adopted in the log–likelihood of the proposed model as described below.

The original log-likelihood function is of the two sub-populations in the mixture model; one for non-negative integers and the other for non-positive integers. While maintaining the mixture of two sub populations in proposed model, the sum of two components in the expression (7.2) is rearranged to a sum of three to accommodate positives, negatives and zeros of the discrete stock price changes . Expression (7.3) shows the three sum of the original two sum log-likelihood given in (7.2).

$$l(\alpha_0, \alpha_1, \beta_0^-, \beta_1^-, \beta_0^+, \beta_1^+) = \sum_{i=1}^{n} \ln \left\{ (1-p_i)P(y_i)I_{y_i \leq 0} + p_i P(y_i)I_{y_i \geq 0} \right\} \qquad (7.2)$$

$$l(\alpha_0, \alpha_1, \beta_0^-, \beta_1^-, \beta_0^+, \beta_1^+) = \sum_{i=1}^{n} \ln \left\{ (1-p_i) \frac{(\lambda_i^-)^{-y_i} e^{-\lambda_i^-}}{-y_i!} \right\} I_{y_i < 0}$$

$$+ \sum_{i=1}^{n} \ln \left\{ p_i \frac{(\lambda_i^+)^{y_i} e^{-\lambda_i^+}}{y_i!} \right\} I_{y_i > 0}$$

$$+ \sum_{i=1}^{n} \ln \left\{ (1-p_i) e^{-\lambda_i^-} + p_i e^{-\lambda_i^+} \right\} I_{y_i = 0} \qquad (7.3)$$

where,

$$p_i = \frac{e^{\alpha_0 + \alpha_1 x_i}}{1 + e^{\alpha_0 + \alpha_1 x_i}}$$

$$1 - p_i = \frac{1}{1 + e^{\alpha_0 + \alpha_1 x_i}},$$

$$\lambda_i^- = e^{\beta_0^- + \beta_1^- x_i} \text{ and}$$

$$\lambda_i^+ = e^{\beta_0^+ + \beta_1^+ x_i}.$$

The equation 7.3 can be further simplified as shown below.

$$l(\alpha_0, \alpha_1, \beta_0^-, \beta_1^-, \beta_0^+, \beta_1^+) = \sum_{i=1}^{n} \left\{ \ln(1-p_i) - y_i(\beta_0^- + \beta_1^- x_i) - \lambda_i^- \right\} I_{y_i < 0}$$

$$+ \sum_{i=1}^{n} \left\{ \ln(1-p_i) + (\alpha_0 + \alpha_1 x_i) + y_i(\beta_0^+ + \beta_1^+ x_i) - \lambda_i^+ \right\} I_{y_i > 0}$$

$$+ \sum_{i=1}^{n} \left\{ \ln(1-p_i) + \ln(e^{-\lambda_i^-} + e^{\alpha_0 + \alpha_1 x_i} e^{-\lambda_i^+}) \right\} I_{y_i = 0}$$

$$- \sum_{i=1}^{n} \left\{ \ln(-y_i!) + \ln(y_i!) \right\} I_{y_i = 0} \qquad (7.4)$$

As given in expression (7.1), 'Delta Method' requires first derivatives of the function interested $(g(\theta))$ and observed information function $J(\theta)$. For the log–likelihood function $l(\theta)$, for $\theta = (\alpha_0, \alpha_1, \beta_0^-, \beta_1^-, \beta_0^+, \beta_1^+)$, the observed information matrix can be expressed as given in expression (7.5). Based on six parameters, the observed information matrix is symmetric and with the order 6 by 6. Expression (7.5) shows the lower triangle of the observed information matrix due its symmetry.

101

$$J(\theta) = \begin{pmatrix} -\frac{\partial^2 l}{\partial \alpha_0^2} & \cdot & \cdot & \cdot & \cdot & \cdot \\ -\frac{\partial^2 l}{\partial \alpha_0 \partial \alpha_1} & -\frac{\partial^2 l}{\partial \alpha_1^2} & \cdot & \cdot & \cdot & \cdot \\ -\frac{\partial^2 l}{\partial \beta_0^- \partial \alpha_0} & -\frac{\partial^2 l}{\partial \beta_0^- \partial \alpha_1} & -\frac{\partial^2 l}{\partial \beta_0^{-2}} & \cdot & \cdot & \cdot \\ -\frac{\partial^2 l}{\partial \beta_1^- \partial \alpha_0} & -\frac{\partial^2 l}{\partial \beta_1^- \partial \alpha_1} & -\frac{\partial^2 l}{\partial \beta_1^- \partial \beta_0^-} & -\frac{\partial^2 l}{\partial \beta_1^{-2}} & \cdot & \cdot \\ -\frac{\partial^2 l}{\partial \beta_0^+ \partial \alpha_0} & -\frac{\partial^2 l}{\partial \beta_0^+ \partial \alpha_1} & -\frac{\partial^2 l}{\partial \beta_0^+ \partial \beta_0^-} & -\frac{\partial^2 l}{\partial \beta_0^+ \partial \beta_1^-} & -\frac{\partial^2 l}{\partial \beta_0^{+2}} & \cdot \\ -\frac{\partial^2 l}{\partial \beta_1^+ \partial \alpha_0} & -\frac{\partial^2 l}{\partial \beta_1^+ \partial \alpha_1} & -\frac{\partial^2 l}{\partial \beta_1^+ \partial \beta_0^-} & -\frac{\partial^2 l}{\partial \beta_1^+ \partial \beta_1^-} & -\frac{\partial^2 l}{\partial \beta_1^+ \partial \beta_0^+} & -\frac{\partial^2 l}{\partial \beta_1^{+2}} \end{pmatrix} \tag{7.5}$$

All the required first and second order derivatives of the likelihood function $l(\theta)$, for $\theta = (\alpha_0, \alpha_1, \beta_0^-, \beta_1^-, \beta_0^+, \beta_1^+)$ are given Appendix C. It should be noted that a significant challenge was faced in computing the derivatives and arranging them based on a pattern.

Section 7.3 presents several approximate confidence intervals based on 'Delta Method'. The function $g(\theta)$ will be changed based on the parameter for the desired confidence interval. θ consists of six population parameters $\alpha_0, \alpha_1, \beta_0^-, \beta_1^-, \beta_0^+$ and β_1^+. Therefore, the derivative of $g(\theta)$, denoted by $\frac{\partial g(\hat{\theta})}{\partial \theta}$, consists of six partial derivatives of each of the six parameters as given in the expression (7.6).

$$\frac{\partial g(\hat{\theta})}{\partial \theta} = \begin{pmatrix} \frac{\partial g(\hat{\theta})}{\partial \alpha_0} \\ \frac{\partial g(\hat{\theta})}{\partial \alpha_1} \\ \frac{\partial g(\hat{\theta})}{\partial \beta_0^-} \\ \frac{\partial g(\hat{\theta})}{\partial \beta_1^-} \\ \frac{\partial g(\hat{\theta})}{\partial \beta_0^+} \\ \frac{\partial g(\hat{\theta})}{\partial \beta_1^+} \end{pmatrix} \tag{7.6}$$

7.3   Approximate CI

7.3.1   Approximate CI for α_1

The importance of the variable mixture probability in the proposed mixture model was highlighted throughout several chapters. In order to variable mixture probability to exist the parameter α_1 should exist. Therefore, the approximate confidence interval for α_1 is examined based on both simulations and the data from FDX.

The 'Delta Method' described in section 7.1 and the observed information matrix $J(\theta)$ given in 7.3 with $g(\hat{\theta}) = \hat{\alpha}_1$ were used to generate the confidence interval.

The data sets for simulations were generated using the true parameters $\alpha_0 = 0.3$, $\alpha_1 = 0.8$, $\beta_0^+ = -0.5$, $\beta_1^+ = 0.2$, $\beta_0^- = -0.7$, and $\beta_1^- = -0.1$. Figure 7.1 shows the plot of 1000 confidence intervals based on the simulated data. It can be seen that majority of the confidence intervals contain the true parameter value, while few excluding the parameter. The computations show that 948 of the 1000 confidence intervals contained the true parameter while 52 excluding the true parameter α_1=0.8. That is closer to the expected 95%. The average 95% confidence interval is (0.7945263, 0.8053108).

Then the 95% confidence interval for α_1 was computed on year 1 FDX data. The data analysis for FDX data of year 1 showed that the estimated value for α_1 is 0.002071554 with the 95% approximated confidence interval (0.001955386, 0.002187722).

Approximate Confidence Interval for $\hat{\alpha}_1$
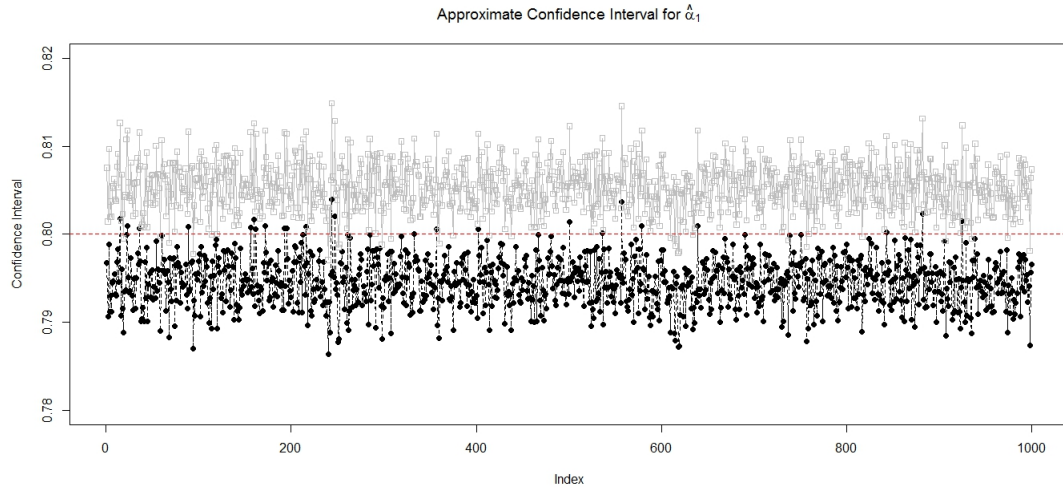
FIGURE 7.1 – Approximate Confidence Interval for α_1. Horizontal line denotes the true value of the parameter, α_1=0.8.

7.3.2 Approximate CI for Variable Mixture Probability

The 95% confidence interval for variable mixture probability (p_i) is computed using the 'Delta Method' described in section 7.1 and the observed information matrix $J(\theta)$ given in 7.3 with $g(\hat{\theta}) = \hat{p}_i$.

$$p_i = \frac{e^{\alpha_0 + \alpha_1 x_i}}{1 + e^{\alpha_0 + \alpha_1 x_i}} \tag{7.7}$$

Figure 7.2 show the approximate confidence interval for variable mixture probability (p_i) on year 1 FDX data.
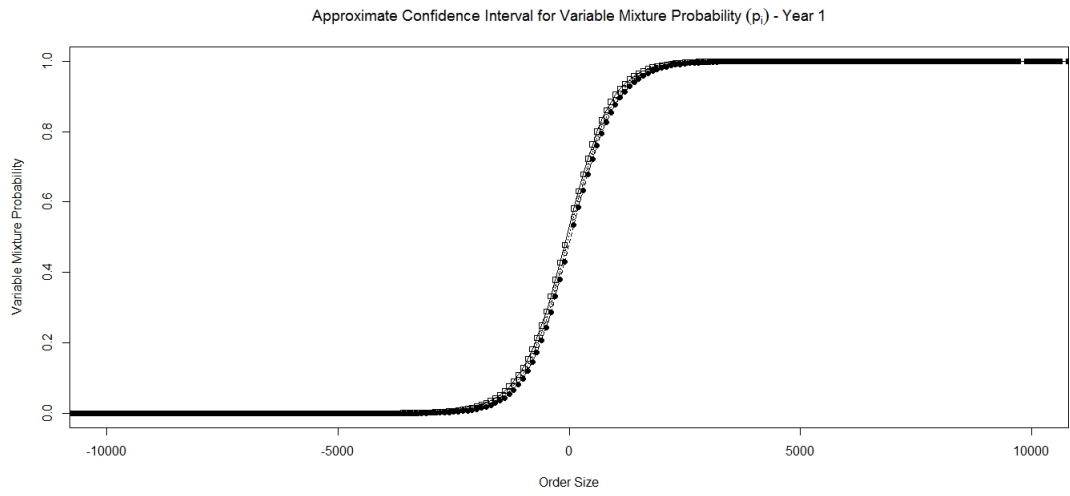
FIGURE 7.2 – Approximate Confidence Interval for Variable Mixture Probability (p_i) on Year 1 FDX data
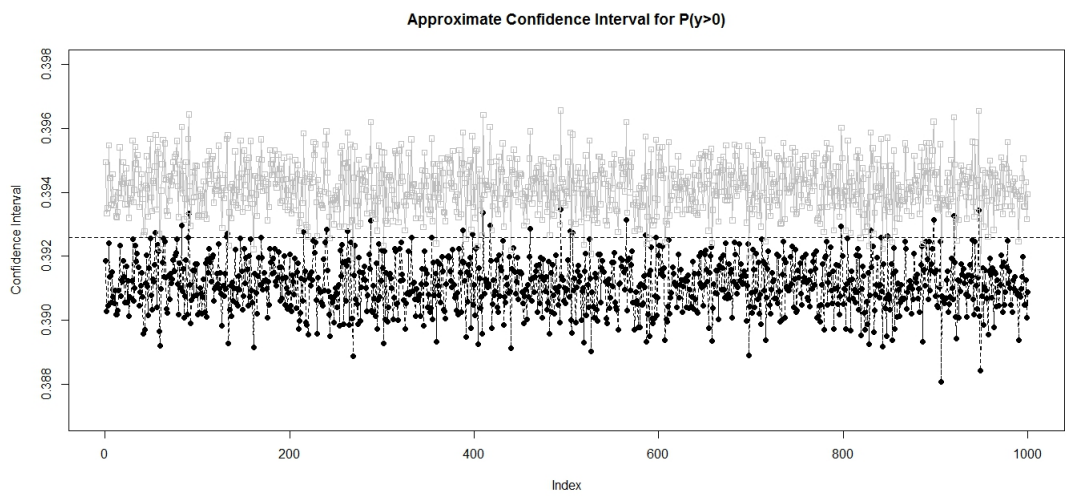
FIGURE 7.3 – Approximate Confidence Interval for $P(y > 0)$ with simulated data

7.3.3 Approximate CI for Probability of Price Change

Three interesting probabilities of the proposed model were discussed in section 3.3.2 of chapter 3, as shown in expressions (7.6), (7.7) and (7.8). They were the probability of the stock price increment ($P(y > 0)$), the probability of the stock price decrement ($P(y < 0)$) and the probability of no price change ($P(y = 0)$).

A similar 'Delta Method' was used in computing the confidence intervals for $P(y > 0)$, $P(y < 0)$ and $P(y = 0)$. $g(\hat{\theta})$ in the expression (7.1) of the method is replaced with $P(y > 0)$, $P(y < 0)$ and $P(y = 0)$ accordingly to find each probability, as given in expressions (7.7), (7.8) and (7.9).

$$P(y_i > 0) = p_i(1 - e^{-\lambda_i^+}) \tag{7.8}$$

$$P(y_i < 0) = (1 - p_i)(1 - e^{-\lambda_i^-}) \tag{7.9}$$

$$P(y_i = 0) = (1 - p_i)e^{-\lambda_i^-} + p_i e^{-\lambda_i^+} \tag{7.10}$$

Figures 7.4, 7.6 and 7.8 show the confidence bands for year 1 for probability of the stock price increment ($P(y > 0)$), the probability of the stock price decrement ($P(y < 0)$) and the probability no price change ($P(y = 0)$) respectively. Figures 7.5, 7.7. and 7.9 show a subsection of the figures 7.4, 7.6 and 7.8 respectively, for closer analysis of the probabilities between order sizes -100 and 100. Figures 7.5, 7.7. and 7.9 how the probabilities of the order sizes change from negative order sizes (sales) to positive order sizes (purchases). It is important to identify that the order size is a discrete variable and does not include 0.

The preliminary analysis of data showed that the order sizes ±100 and ±200 occur with a significantly large frequency. The thick confidence band closer to 0 in the figures 7.4, 7.6 and 7.8 are due to those high frequent transaction of order sizes ±100 and ±200.

According to the figure 7.4, the probability of stock price increment becomes higher for purchases with larger order sizes. Confidence interval is larger for less
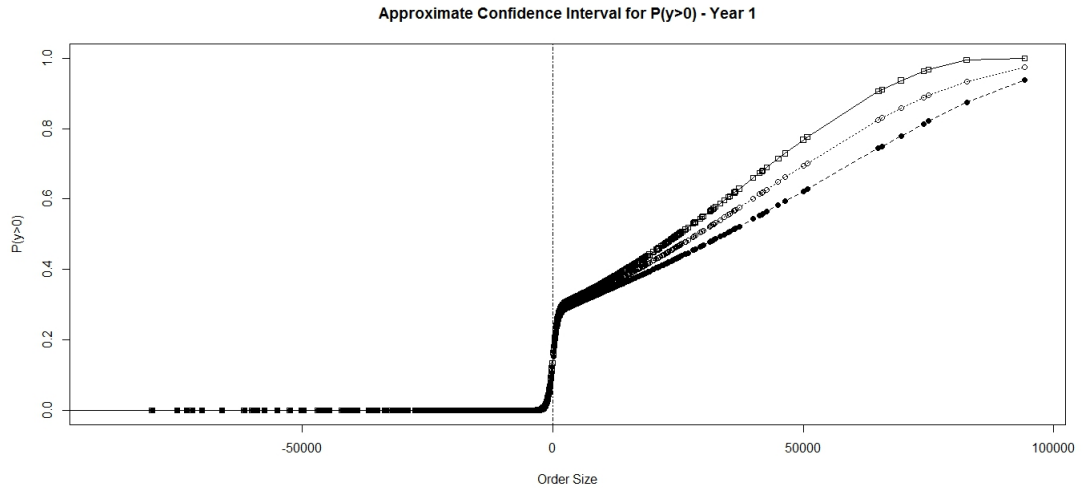
FIGURE 7.4 – Approximate Confidence Interval for $P(y > 0)$ on Year 1 FDX data

frequent and smaller for more frequent order sizes. The probability of stock price decrement becomes higher for larger sales, as seen in figure 7.6. Similarly, the confidence interval is larger for less frequent sales. Figure 7.8 shows that the smaller order sizes have more probability towards keeping the stock price stable than larger order sizes. The confidence intervals of the smaller order sizes are smaller than the larger order sizes.

A jump in the probability of no price change ($P(y = 0)$) can be seen between the order sizes -100 and 100 in figure 7.9. According to year 1 FDX data, the negative order sizes have more impact on changing the stock price than the positive order sizes. Moreover, sales of stocks have more probability to change the stock price than purchases. The figures further show that the proposed mixture model conforms the expectations of the stock market on the probability of stock price change.
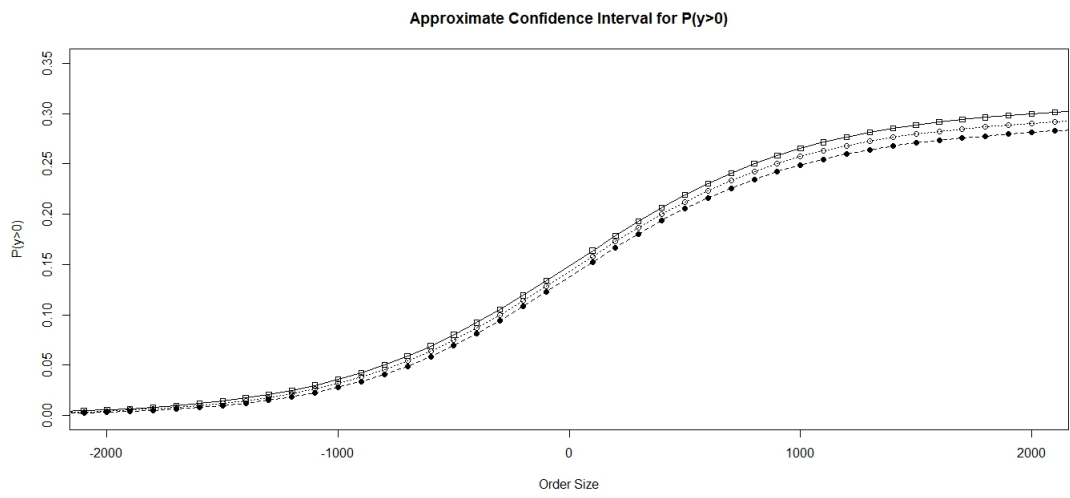
FIGURE 7.5 – A sub section of figure 7.4
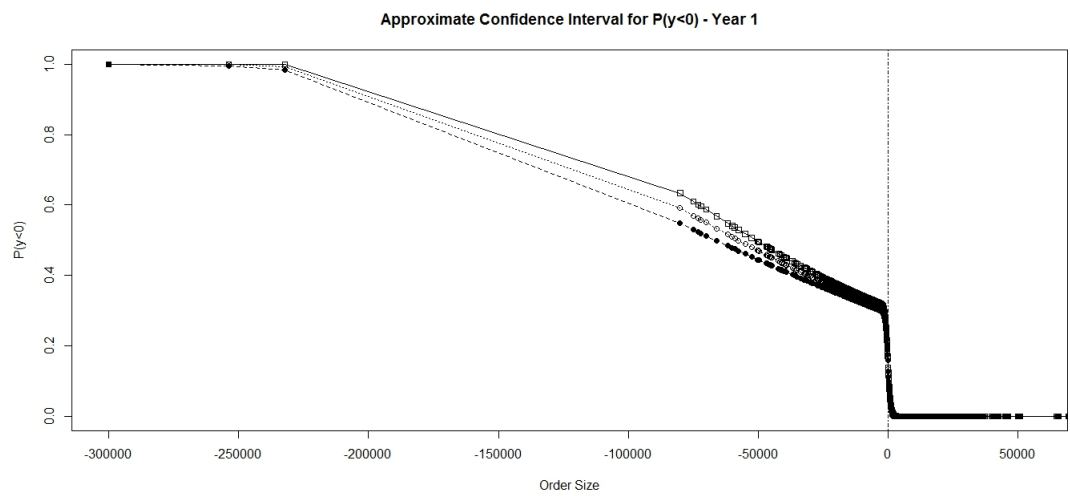
FIGURE 7.6 – Approximate Confidence Interval for $P(y < 0)$ on Year 1 FDX data
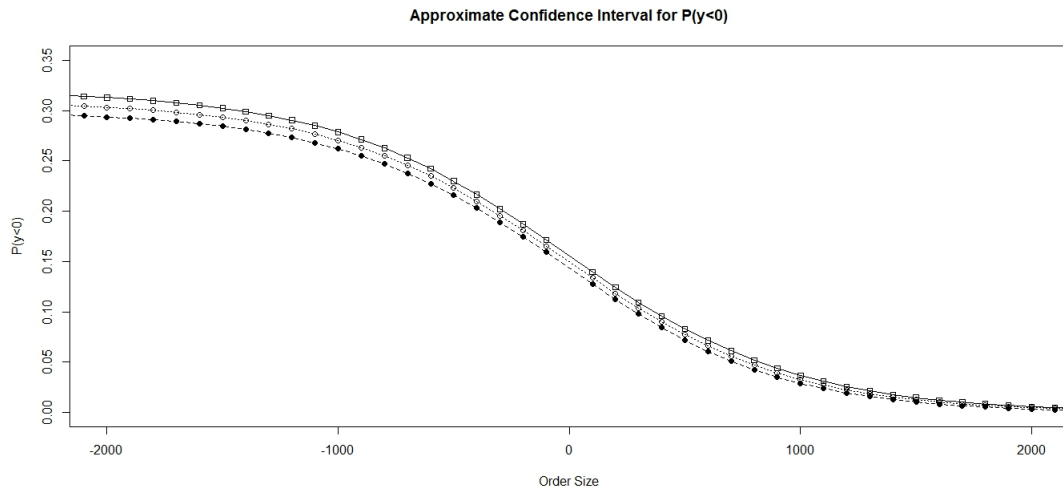
FIGURE 7.7 – A sub section of figure 7.6
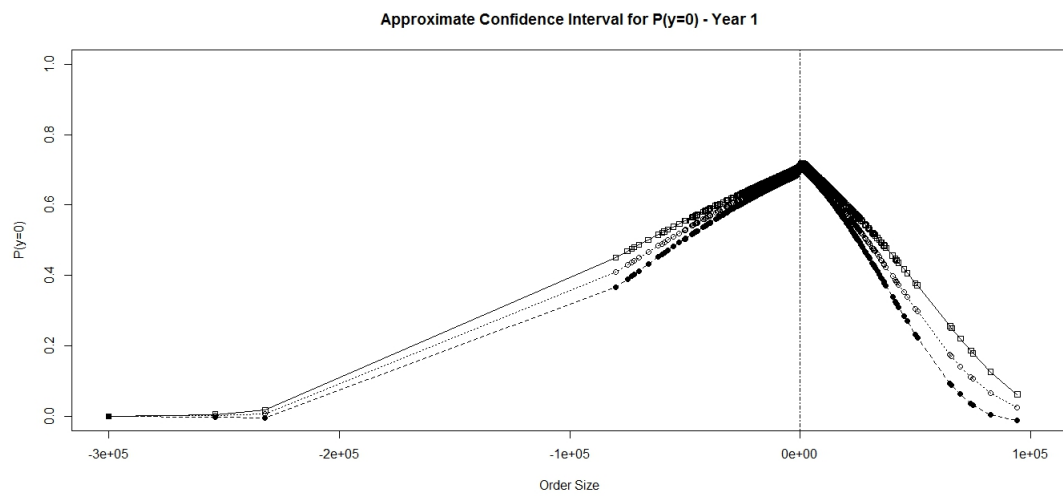
FIGURE 7.8 – Approximate Confidence Interval for $P(y = 0)$ on Year 1 FDX data

Approximate Confidence Interval for P(y=0)
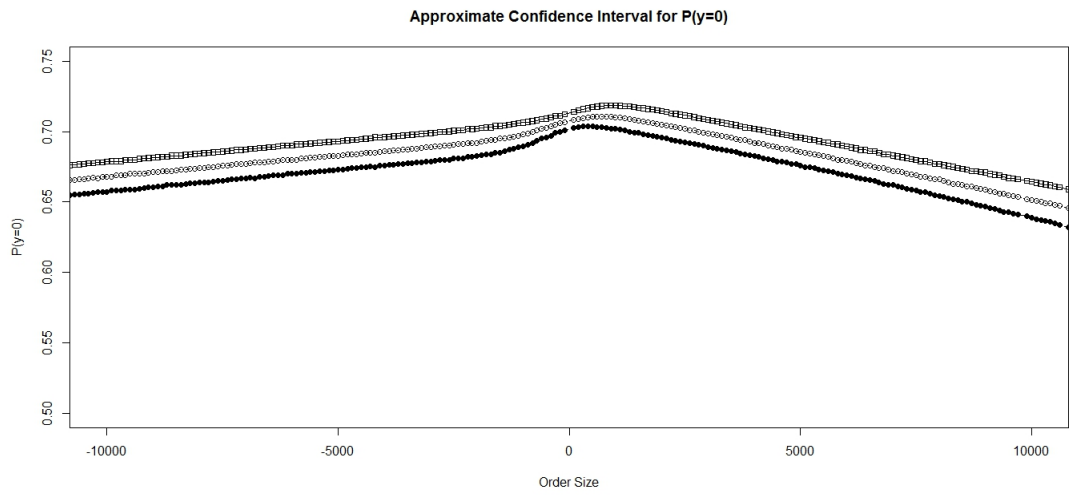
FIGURE 7.9 – A sub section of figure 7.8

CHAPTER 8
DISCUSSION AND CONCLUSION

A novel method of modeling stock price changes using a mixture model was proposed under the research performed on tick-by-tick stock transactions data. The stock price changes were analyzed based on the minimum price movement known as 'tick-size'. The most natural distribution for discrete data, the Poisson distribution, was used to model the discrete stock price changes. The model was proposed based on a constant mixture probability and also with a variable mixture probability which depends on the order size.

Maximum likelihood method was used to estimate the model parameters with the use of Expectation-Maximization (EM) algorithm. The model was evaluated using simulated data with known parameters. The results were acceptable and it was identified that the estimates converge to the true parameters as the size of the data sets were increased. Tick-by-tick stock transactions from Federal Express, were analyzed with the proposed model. Three interesting probabilities of stock price change, namely, probability of the stock price increment ($P(y > 0)$), the probability of the stock price decrement ($P(y < 0)$) and the probability of no price change ($P(y = 0)$) were also computed based on the proposed model.

The proposed model was implemented using the statistical programming language **R**. As a resolution for the challenge of efficiency, the implementations were adjusted with user written codes and also implementing one of the most recent versions of the EM algorithm, which is know as 'PEM'. Further the university HPC cluster was utilized for parallel processing of the model. As another resolu-

tion for speeding up the model, a 'Clustered Signed Model' was proposed to using summarized data to reduce the amount of data to be used in the model implementation. The discreteness of the order size and the sign of the discrete stock price change were used. The clustered model exhibited a significant gain in time compared the method discussed under the efficiency improvements.

A parametric bootstrap procedure was considered to assess the significance of the order size on the mixing probabilities. The results of the parametric bootstrap shows that the use of a variable mixture probability, which depends on the order size, is more appropriate for the model, as the stock price changes do depend on the order size. The methods are illustrated with data from simulations and real data from Federal Express.

8.1   Model Consequences

There are several significant consequences of the proposed mixture model of two poisson distributions.

- Novelty :

  There have been a large amount of research performed on stock transactions data treating the stock prices as continuous values. Based on the stock market regulated 'tick-size' the proposed model treats the stock price changes as a set of discrete values. The discreteness of the stock price changes is clearly adds a novelty to the proposed model.

- Variable Mixture Probabilities (p_i) :

  The use of variable mixture probability as a function of order size, in the mixture model can also be highlighted as novelty in the model, where the mixture probability p_i in the equation (8.1).

$$p_i = \frac{e^{\alpha_0 + \alpha_1 x_i}}{1 + \alpha_0 + \alpha_1 x_i} \tag{8.1}$$

The literature suggests that the most common way of handling the mixture probability is as treating the mixture probability as a constant. Order size highly influences the change of the stock price. Therefore, making the variable mixing probability to be a function of order size seemed to be a more realistic assumption and later it was demonstrated to be more appropriate on read data using the parametric bootstrap method.

- Different Parametric Formulation :

  The use of Mixtures of Poisson distributions are found in many different applications. The most common setting is, when handling two sub populations of **non-negative** integers in a population. That uses a similar and standard parametric formulation of the poisson distribution as given in equations (8.2).

  $$Y_1 \sim Poisson(\lambda_1) \tag{8.2}$$
  $$Y_2 \sim Poisson(\lambda_2)$$
  $$\text{where } Y_1 = Y_2 = \{0, 1, 2, \dots\}$$

  The mixture of discrete stock price changes consists of both stock price increments and decrements. Therefore, the population under research is composed of a mixture of a set of **non-negative** integers which was resulted from stock price increments and a set of **non-positive** integers which was resulted from stock price decrements. The non-negative integers clearly follow a poisson distribution, however, the non-positive integers were needed to be negated to apply a poisson distribution. This results a different parametric formulation of the poisson distribution to be used in the mixture model as shown in the equation (8.3).

$$Y_1 \sim Poisson(\lambda^+) \qquad (8.3)$$

$$-Y_0 \sim Poisson(\lambda^-)$$

where $Y_1 = \{0, 1, 2, \dots\}$ and $Y_0 = \{\dots, -3, -2, -1, 0\}$

With the common mixture models of poisson or normal distributions, both the mixtures contain data with the same range of values. For example, if in a mixture of normal distributions, each sub population contains continuous real values. Then EM does not identify whether the group of estimates are from the first sub population or the second sub population. In such cases, there is a considerable possibility of iterating the estimates between the two sub populations and making the convergence of EM more difficult.

Whereas, in the proposed model, the mean parameter of the non-negative integers is positive and the mean parameter of the non-positive integers is negative. When estimating parameters of the model, EM algorithm has an additional knowledge as one sub population has non-negative values and the other sub population has non-positive values. That will make the estimation more convenient for the EM algorithm, adding an advantage when performing parameter estimation in the proposed model.

- Clustered Signed Model :
  The initially proposed mixture model consumes the processed data of tick-by-tick transactions. The processed data includes the discrete stock price change and its corresponding order size. There is a large amount of data with discrete stock price change and its order size. Although, there were two variables used for the model (y_i and x_i), the amount of data are still massive. Under the 'Clustered Signed Model' the proposed model was further

114

simplified using the known properties of the two variables. Both the order size and stock price changes were discrete variables, and the orders are placed in multiples of hundreds and hundred being the most frequent value.

The summarized values include clustered x_i values, the number of y_i values that are zeros (N_{i0}), the number of y_i values that are negative (N_{i-}), the number of y_i values that are positive (N_{i+}), the sum of positive y_i values (y_{i+}) and the absolute sum of negative y_i (y_{i-}). An additional functionality was also implemented to summarize the data set of order size (x_i) and the discrete stock price changes (y_i) as above, in order to use with the clustered model. Simulations show that there is a significant gain in the efficiency from the proposed clustered model. The constant model of the clustered model even outperforms the implementation of the PEM algorithm of the clustered model.

- Approximate Confidence Interval :

  One of the most popular method of finding confidence intervals, 'Delta Method' was used to estimating the approximate confidence intervals of the the parameters. To avoid the complexity of differentiation, a trick was employed to divide the original log-likelihood of two sum of two mixtures, was changed to be a three sum of positives, negatives and also zeros of the discrete stock price changes.

  Approximate confidence intervals were computed for α_1, the variable mixture probability (p_i), probability of stock price increment $(P(y > 0))$, the probability of the stock price decrement $(P(y < 0))$ and the probability no price change $(P(y = 0))$.

8.2   Future Research Directions

The research performed on the proposed model and the tick-by-tick stock transactions data offer many interesting paths to continue investigations. A few are described below.

1. Asymptotic Properties

    There has been limited theoretical research performed on theoretical asymptotic properties of the estimates of the parameters in mixture models. Therefore, it would be interesting to find a way to theoretically derive the asymptotic distributions, weak and strong consistencies of the parameter estimates. The work of Fahrmeir and Kaufmann (1985) on the consistency and asymptotic normality of the Maximum Likelihood Estimator in Generalized Linear Models has made an important milestone, thus provides support to prove asymptotic properties of the estimates of the parameters in the proposed mixture models.

2. Recover the Supply Chain Curve

    Cetin *et al.* (2006) assumes the stock's supply curve satisfies the equation given in

    $$S(t, x) = e^{\alpha x} S(t, 0) \text{ with } \alpha > 0 \tag{8.4}$$

    where

    $$S(t, 0) = \frac{s_0 exp\{\mu t + \sigma W_t\}}{exp\{rt\}} \tag{8.5}$$

    for constants μ and σ with W_t denoting a standard Brownian motion and spot rate of interest r. $S(t, x)$ represents the stock price, per share, at time $t \in [0, \text{T}]$ that a trader pays/receives for order flow x normalized by the value of a money market account as described by Cetin *et al.* in 2006.

    It would be interesting to work on recovering the supply curve $S(t, x)$ from the stock price used in the Poisson mixture model. Preliminary progress was

made for the case when is close to 0 and assume $S(t_{i-1}, x_{t_{i-1}}) = S(t_{i-1}, 0)$ to avoid the confounded effects between the previous order size and the current order size.

3. <u>Time of the Transaction</u>

The covariates of the proposed mixture model only the order size. The model seemed reasonable with order size due to the fact that the size of a purchase order will increase the stock price more and the size of a sell order will decrease the stock price more. However, it is an interesting question to investigate that whether the time of transaction within the day has an effect towards the stock price change.

During one of the earlier investigations based on an extension of Gill *et al.* (2007), on "Multiple change point analysis on stock transactions" it was identified that the stock price has significant changes during the beginning, middle and end of the day. Therefore, 'time of day' is also an important factor when it comes to the volatility of the stock price. With that, the next step is to investigate ways to incorporate the 'time' into the Poisson mixture model.

REFERENCES

Bauerle, N., and U. Rieder. 2004. Portfolio optimization with Markov-modulated stock prices and interest rates. Automatic Control, IEEE Transactions on 49: 442-447.

Berlinet, A. and C. Roland. 2009. Parabolic acceleration of the EM algorithm. Statistics and Computing 19: 35-47.

Berlinet, A. F. and C. Roland. 2012. Acceleration of the EM algorithm: P-EM versus epsilon algorithm. Computational Statistics & Data Analysis 56: 4122-4137.

Blekas, K., Likas, A., Galatsanos, N. P. and I. E. Lagaris. 2005. A Spatially Constrained Mixture Model for Image Segmentation. IEEE Transactions of Neural Networks 16: 494 - 498.

Bollen, J., H. Mao, and X. Zeng. 2011. Twitter mood predicts the stock market. Journal of Computational Science 2: 1-8.

Bonate, P. L. 2011. *Pharmacokinetic-pharmacodynamic Modeling and Simulation.* 2nd edn. Springer New York.

Brame, R., D. S. Nagin, and L. Wasserman. 2006. Exploring Some Analytical Characteristics of Finite Mixture Models. Journal of Quantitative Criminology 22: 31-59.

Brijs, T., Kharlis, D., Swinnen, G., Vanhoot, K., Wets, G. and P. Manchanda. 2004. A multivariate Poisson mixture model for marketing applications. Statistica Neerlandica 58: 322-348.

Caudill, S., Gropper, D. and V. Hartarska. 2009. Which Microfinance Institutions

Are Becoming More Cost Effective with Time? : Evidence from a Mixture Model. Journal of Money, Credit and Banking 41: 651 - 672.

Cetin, U., Jarrow, R., Protter, P. and M. Warachka. 2006, Pricing options in an extended Black Scholes economy with illiquidity: theory and empirical evidence. Rev. Financial Stud. 19, 493529.

Chen, J. 1995. Optimal Rate of Convergence for Finite Mixture Models. The Annals of Statistics 23: 221-233.

Chen, T.-L., C.-H. Cheng, and H. Jong Teoh. 2007. Fuzzy time-series based on Fibonacci sequence for stock price forecasting. Physica A: Statistical Mechanics and its Applications 380: 377-390.

Chernick, M. R. 1999, *Bootstrap Methods : A Practitioner's Guide*, Wiley Series.

Chung, F., Fu, T., Luk, R. and V. Ng. 2002. Evolutionary time series segmentation for stock data mining. In: Data Mining, 2002. ICDM 2003. Proceedings. 2002 IEEE International Conference on. p 83-90.

Czado, C. and A. Min. 2005. Consistency and Asymptotic Normality of the Maximum Likelihood Estimator in a Zero–Inflated Generalized Poisson Regression. Technical Paper. <http://epub.ub.uni-muenchen.de/1792/1/paper_423.pdf >. [ACCESS DATE] May 2013.

Davison, A. C. and Hinkley, D. V, 1997, *Bootstrap Methods and Their Applications*, Cambridge University Press.

De Adana, F.S., Gutierrez, O. and I. Gonzalez. 2011. *Practical Applications of Asymptotic Techniques in Electromagnetics.*

Dempster, A. P., Laird, N. M. and D. B. Rubin. 1977, Maximum likelihood from incomplete data via the EM algorithm, J. R. Statist. Soc, 39, 1-3

Diaz, D., B. Theodoulidis, and P. Sampaio. 2011. Analysis of stock market manipulations using knowledge discovery techniques applied to intraday trade prices. Expert Systems with Applications 38: 12757-12771.

Dunteman, G. H. 2004. *An Introduction to Generalized Linear Models.* SAGE
Publications.

Enke, D., and S. Thawornwong. 2005. The use of data mining and neural networks
for forecasting stock market returns. Expert Systems with
Applications 29: 927-940.

Fahrmeir, L. and H. Kaufmann. 1985. Consistency and Asymptotic Normality of
the Maximum Likelihood Estimator in Generalized Linear Models. The
Annals of Statistics 13: 342-368.

Feng, W., Liu, Y., Wu, J., Nephew, Huang, T. and L. Li. 2008. A Poisson
mixture model to identify changes in RNA polymerase II binding quantity
using high-throughput sequencing technology. BMC genomics 9 Suppl 2:
S23.

Finch, S. J., Mendell, N. R. and H. C. Thode JR. 1989. Probabilistic Measures
of Adequact if a Numerical Search for Global Maximum. Journal of
American Statistical Association 84: 1020-1023.

Garthwaite, P. H., Jolliffe, I. T. and B. Jones. 2002. *Statistical Inference.* 2nd
edn. Oxford Science Publications.

Gavrishchaka, V. and S. Banerjee. 2006. Support Vector Machine as an Efficient
Framework for Stock Market Volatility Forecasting. Computational
Management Science 3: 147-160.

Gerdtham, G. and P. K. Trivedi. 2001. Equity in Swedish Health Care
Reconsidered : New Results Based on the Finite Mixture Model. Health
Econ. 10: 565572.

Gill, R., K. Lee, and S. Song. 2007. Computation of estimates in segmented
regression and a liquidity effect model. Computational Statistics & Data
Analysis 51: 6459-6475.

Greenspan, H., Goldberger, J. and I. Eshet. 2001. Mixture Model for Face-color

Modeling and Segmentation. Pattern Recognition Letters 22: 1525 - 1536.

Guo, C., Jia, H., and N. Zhang. 2008. Time Series Clustering Based on ICA for Stock Data Analysis. In: Wireless Communications, Networking and Mobile Computing, 2008. WiCOM '08. 4th International Conference on. p 1-4.

Harris, L. 1991. Stock price Clustering and Discreteness. The Review of Financial Studies 4: 389-415.

Harper, D. 2013, Forces That Move Stock Prices <http://www.investopedia.com/ articles/basics/04/100804.asp >, [ACCESS DATE] 23 April 2013.

Hernandez, J. A. and I. W. Phillips. 2006. Weibull mixture model to characterize end-to-end Internet delay at coarse time-scales. IEE Proc.-Communs.153.

History of FedEx Operating Companies. 2013. <http://about.van.fedex.com/fedex-opco-history >, [ACCESS DATE] 23 April 2013.

Hsu, Y.-T., M.-C. Liu, J. Yeh, and H.-F. Hung. 2009. Forecasting the turning time of stock market based on MarkovFourier grey model. Expert Systems with Applications 36: 8597-8603.

Huang, K. Y., and C.-J. Jane. 2009. A hybrid model for stock market forecasting and portfolio selection based on ARX, grey system and RS theories. Expert Systems with Applications 36: 5387-5392.

Huang, W., Y. Nakamori, and S.-Y. Wang. 2005. Forecasting stock market movement direction with support vector machine. Computers & Operations Research 32: 2513-2522.

Jilani, T. A., and S. M. A. Burney. 2008. A refined fuzzy time series model for stock market forecasting. Physica A: Statistical Mechanics and its Applications 387: 2857-2862.

Karlis, D. and E. Xekalaki. 2003. Choosing Initial Values for the EM Algorithm for Finite Mixtures. Computational Statistics & Data Analysis 41: 577-590.

Li, F., and C. Liu. 2009. Application Study of BP Neural Network on Stock
Market Prediction. In: Hybrid Intelligent Systems, 2009. HIS '09. Ninth
International Conference on. p 174-178.

Madsen, H. and P. Thyregod. 2011. *Introduction to General and Generalized
Linear Models.* CRC Press.

Martnez-Camblor, P and N. Corral. 2012. A general bootstrap algorithm for
hypothesis testing, Journal of Statistical Planning and Inference 142 : 2,
589-600.

McLachlan, G. and D. Peel. 2000. *Finite Mixture Models.* Wiley Series.

Moro, E. Vicente, J. Moyano, L. Gerig, A. Farmer, J. Vaglica, G. Lillo, F. and
R. Mantegna. 2009. Market impact and trading profile of hidden orders in
stock markets. Physical Review E 80: 066102.

Moroney, R. 2007. Seven Stock Split Superstars <http://www.forbes.com/2007/
02/01/fedex-lockheed-aflac-pf-guru-in_rm_0201soapbox_inl.html >,
[ACCESS DATE] 21 November 2012.

Oh, K. J., and K.-j. Kim. 2002. Analyzing stock market tick data using piecewise
nonlinear model. Expert Systems with Applications 22: 249-255.

O'Hagan, A., T. B. Murphy, and I. C. Gormley. 2012. Computational aspects of
fitting mixture models via the expectation maximization algorithm.
Computational Statistics & Data Analysis 56: 3843-3864.

Park, B. J., and D. Lord. 2009. Application of finite mixture models for vehicle
crash data analysis. Accident; analysis and prevention 41: 683-691.

Pawitan, Y. 2001. *In All Likelihood Statistical Modeling and Inference Using
Likelihood.* Oxford University Press, Oxford.

PET. 2011. <http://www.webmd.com/a-to-z-guides/positron-emission-tomography
>. [ACCESS DATE] 13 April 2013.

Plerou, V., P. Gopikrishnan, X. Gabaix, and H. E. Stanley. 2002. Quantifying

stock-price response to demand fluctuations. Physical Review E 66: 027104.

Podobnik, B., D. Horvatic, A. M. Petersen, and H. E. Stanley. 2009. Cross-correlations between volume change and price change. Proceedings of the National Academy of Sciences 106: 22079-22084.

Rocha, M., Soldevilla, M. and B. Burtenshaw. 2007. Gaussian mixture model classification of odontocetes in the Southern California Bight and the Gulf of California. Acoustical Society of America 121.

Scalas, E. 2007. Mixtures of compound Poisson processes as models of tick-by-tick financial data. Chaos, Solitons & Fractals 34: 33-40.

Shalizi, C. 2011, The Bootstrap. <http://www.stat.cmu.edu/ cshalizi/402/lectures/ 08-bootstrap/lecture-08.pdf>. [ACCESS DATE] 12 May 2013.

Stock Market. 2013. <http://www.investopedia.com/terms/s/stockmarket.asp >. [ACCESS DATE] 15 April 2013.

Stock Market History. 2005. <http://www.money-zine.com/investing/stocks/ stock-market-history/>. [ACCESS DATE] 15 April 2013.

Su, K. Chen, J. Lee, J. Hu, C. Chang, C. Chou, Y. Liu, R and J. Chen. 2011. Image segmentation and activity estimation for microPET 11C-raclopride images using an expectation-maximum algorithm with a mixture of Poisson distributions. Computerized medical imaging and graphics : the official journal of the Computerized Medical Imaging Society 35: 417-426.

Tick Size. 2013. <http://www.investopedia.com/terms/t/tick-size.asp>. [ACCESS DATE] 3 December 2012.

Uusipaikka, E. 2008. *Confidence Intervals in Generalized Regression Models*. CRC Press.

What is a stock split?, Why do stocks split?. 2009. <http://www.investopedia.com /ask/answers/113.asp>. [ACCESS DATE] 15 April 2013.

Woehrmann, P. 2007. Discrete Stock Prices and Predictability. paper presented

to National Centre of Competence in Research Financial Valuation and Risk Management. July.

Wu, C. F. J. 1983. On the Convergence Properties of the EM Algorithm. The Annals of Statistics 11: 95-103.

Xu, L. and M. Jordan. 1996. On Convergence Properties of the EM Algorithm for Gaussian Mixtures. Neural Computation 8: 129-151.

Yao, J. and S. Kong. 2008. The application of stream data time-series pattern reliance mining in stock market analysis. In: Service Operations and Logistics, and Informatics, 2008. IEEE/SOLI 2008. IEEE International Conference on. p 159-163.

Yao, W. 2013. A note on EM algorithm for mixture models. Statistics & Probability Letters 83: 519-526.

Yu, L., C. Huanhuan, W. Shouyang, and L. Kin Keung. 2009. Evolving Least Squares Support Vector Machines for Stock Market Trend Mining. Evolutionary Computation, IEEE Transactions on 13: 87-102.

Zhai, C., Velivelli, A. and B. Yu. 2004. A Cross Collection Mixture Model for Comparative Text Mining. Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining : 743-748.

Zhou, Y., and Z. Jie. 2010. Stock Data Analysis Based on BP Neural Network. In: Communication Software and Networks, 2010. ICCSN '10. Second International Conference on. p 396-399.

Zhu, W. 1997. Making bootstrap statistical inference : A Tutorial. Research Quarterly for Exercise and Sport 68: 44-55.

APPENDIX A - SAMPLE DATASETS

A.1 Raw Data

The original dataset that was obtained from FDX consisted of daily tick–by–tick transactions. Log–stock price, order size and the time of the transaction was available for each transaction. A sample of the original dataset is given below.

```
logprice ordersize time
4.2626801 2000 34801
4.2609177 -500 35136
4.2591524 -1200 35201
4.2609177 300 35298
4.2609177 500 35868
4.2591524 -2200 37648
4.2573843 -1000 39142
4.2573843 200 40280
4.2556129 -200 40541
4.2556129 -700 40554
4.2556129 200 41528
4.2520604 -3400 42565
```

A.2 Processed Data for the Mixture Model

In order to make the data suitable to use in the proposed mixture model, the discrete stock price change was computed. The following computations were performed on raw data.

1. Log–price exponentiated to obtain the stock price

2. Stock price adjusted according to tick-size

3. Number of ticks between each consecutive transaction computed.

Processed data for the proposed mixture model is shown below. Y denotes the discrete stock price change and X is the order size.

```
Y  X
-1 -500
-1 -1200
1 300
0 500
-1 -2200
-1 -1000
0 200
-1 -200
0 -700
0 200
```

A.3 Processed Data for the Clustered–Signed Model

The Clustered–Signed model also requires processed data. The data used in section A.2 is further processed based on the similarity and the sign of the values as described below.

1. Order size x_i is classified based on similarity of values, denoted by xc.

2. For each distinct x_i, the following are computed.

   (a) compute the sum of positive y_{ij}, denoted by yp.

   (b) compute the sum of negative y_{ij}, denoted by yn.

   (c) compute the number of positive y_{ij}, denoted by Np.

   (d) compute the number of negative y_{ij}, denoted by Nn.

   (e) compute the number of zero y_{ij}, denoted by $N0$.

A sample of data processed for clustered–signed model is shown below.

```
xc :
-500   -1200    300     500  -2200  -1000     200    -200    -700   -3400
-1500   -5000  -1400   -4400   1200 -11000     100    2600    2000    -300
-1300    8300    800    6000   1700   1500   -4800  -19700   -2000     400
-400     700   1600    2400   -800    -900   -4000   -1700    3200    1800


yn :
330   73   14   35   26 460   23 410   95   13    0    0   21   29   78   95 343    0   17
74  115  104   34    4    2    6   10    0    0  181   47    0    2    0    2    5    5    3
189    6  151    2    3    5   69   56   39   19    0    3    4    2   55   27    3    2    0
78    3   15    0    0    3    0    0    0    0    3    1    0    3    4    1    0    0    2


yp :
16    1  200  379    0   34  292    9    8    0   16    6    1  470    0   39   11    0    1
0    3    2    1    0   47    0  365    8  201    5    2    1   79   20   28  111    0    0
4   87   34   12    7    1    0    0   16   28   59   58    0    2  117   73    0    0   43
0    0   17    0   55    3    5    0    0    6    4   91    0    0   10   30   54    0    0


Nn :
287   61    7   15   23  386   13  399   81   10    0    0   20   11   67   83  329    0
14   57   94   89   32    3    2    5    7    0    0  161   37    0    2    0    2    3
4    2  157    5  135    2    2    2   60   52   32   16    0    2    2    2   46   25
2    1    0   59    3   13    0    0    2    0    0    0    0    2    1    0    2    4


Np :
9    1  174  327    0   13  273    9    3    0   13    5    1  405    0    3    9    0
1    0    2    2    1    0   41    0  334    7  174    4    1    1   73   17   26   95
0    0    4  108    2   84   32   11    3    1    0    0   14   24   52   51    0    1
107   65    0    0   37    0    0    4    0   44    2    4    0    0    5    3   78


N0 :
670  105  545  782   44  668 1013 1266  182   20   26    4   74  800
103  271 1150    2   24   83  145  107   60   12   80    4 1145   26
354  559   67    1  179   20   55  170   10    1  253  370  422  167
78   34  139  124   46   53   22   40   78   95  103   73  216  124
```

APPENDIX B - PEM CODE

B.1 R Code for Parabolic-EM Algorithm

The **R** implementation of PEM algorithm (Berlinet and Roland, 2012) is given below. The PEM is implemented in the function 'PEM_Const'. 'PEM_Const' uses three other functions within its functionality; 'like' function, 'Func' function and the 'NR_poisson' function.

'like' function calculates the log-likelihood of the poisson mixture model for given values of x, y and estimates denoted by P. The function 'Func' implements the successive 'E' and 'M' steps of original EM algorithm, that is needed in PEM. 'Func' operates on similar arguments as 'like'. 'NR_poisson' computes the maximum likelihood estimates.

```
# Newton Raphson implementation for poisson regression
NR_poisson=function(X,y,epsilon=.0000001,max.iter=1000,k=1){
        b.new=c(0,0)
        diff.b0=1
        diff.b1=1
        i=1
        while( ((abs(diff.b0)>epsilon) |
                abs(diff.b1)>epsilon))& i<max.iter){
                b.old=b.new
                theta.old=X%*%b.old
                lambda.old=exp(theta.old)
                W=c(lambda.old)
                m=lambda.old
                b.new=b.old+solve(t(k*X)%*%(W*X),
                        t(k*X)%*%(y-m))
        diff.b0=b.old[1]-b.new[1]
        diff.b1=b.old[2]-b.new[2]
        i=i+1
```

```r
        }
        b.new
}




# computes the likelihood
like=function(x, y, P){
        coef.p=P[1:2]
        coef.n=P[3:4]
        pi=P[5]
        lik=sum(log((1-pi)*dpois(-y,lambda=exp(coef.n[1]
                +coef.n[2]*x))+ pi*dpois(y,lambda=
                exp(coef.p[1]+coef.p[2]*x))))
        lik
}




# successive E and M steps needed for PEM
Func=function(x,y, P){
        n=length(y)
        yp=y[y>=0]
        yn=-y[y<=0]
        xp=x[y>=0]
        xn=x[y<=0]
        coef.p=P[1:2]
        coef.n=P[3:4]
        pi=P[5]
        gamma=pi*exp(-exp(coef.p[1]+coef.p[2]*x))/
                ((1-pi)*exp(-exp(coef.n[1]+coef.n[2]*x))
                +pi*exp(-exp(coef.p[1]+coef.p[2]*x)))
        gamma=(y>0)*1+(y<0)*0+gamma*(y==0)
        k=gamma
        kp=k[y>=0]
        kn=1-k[y<=0]
        coef.p=NR_poisson(cbind(1,xp),yp,k=kp)
        coef.n=NR_poisson(cbind(1,xn),yn,k=kn)
        pi=sum(k)/n
        P=rbind(coef.p[1],coef.p[2],coef.n[1],coef.n[2],pi)
}
```

```
# PEM algorithm implementation
PEM_const=function(x,y,param.init=c(rep(0,4),.5),
        epsilon=.0000001,max.iter=10000){
        print(param.init)
        n=length(y)
        # 1-2 positive intercept and slope parameters,
        # 3-4 negative intercept and slope parameters,
        # 5 is p
        P0=rbind(param.init[1],param.init[2],param.init[3]
                ,param.init[4],param.init[5])
        another.step=TRUE
        Lold=like(x, y, P0)
        P1=Func(x,y,P0)
        P2=Func(x,y,P1)
        Pold=P0
        iter=0
        while ((iter<=max.iter)& another.step){
                iter = iter + 1
                Pbest = P2
                Lbest = like(x,y, P2)
                # geometric grid search
                i = 0
                t = 1.1
                Pnew = (0.01*P0)-(0.22*P1)+(1.21*P2)
                Lnew = like(x,y, Pnew)
                while (Lnew > Lbest){
                        Pbest = Pnew
                        Lbest = Lnew
                        i = i + 1
                        t = 1 + ((1.5^i)*0.1)
                        Pnew = ((1-t)*(1-t))*P0+(2*t*(1-t))
                                *P1+(t^2)*P2
                        Lnew = like(x,y, Pnew)
                }
                another.step=(max(abs(Pbest-Pold))>epsilon)
                P0 = P1
                P1 = P2
                P2 = Func(x,y, Func(x,y, Pbest))
                Lold = Lbest
                Pold=Pbest
        }
        if (another.step==TRUE){
                cat("EM algorithm did not converge in ",
                max.iter," iterations\n")
```

```
            }
            coef.p=P2[1:2]
            coef.n=P2[3:4]
            pi=P2[5]
            lik=Lbest
            est=cbind(p0=coef.p[1],p1=coef.p[2],n0=coef.n[1],
                    n1=coef.n[2], p=pi, lik ,count=iter+2)
            est
}
```

APPENDIX C - DERIVATIVES

C.1 First Derivatives

The first derivatives of the log–likelihood function that is given in expression (7.4) in chapter 7 are needed for the approximate confidence interval. The log-likelihood function is differentiated with respect to each of the six parameters of the model. The six first derivatives are given below.

$$
\begin{aligned}
\frac{\partial l}{\partial \alpha_0} &= \sum_{i=1}^{n} \{-p_i I_{y_i<0}\} + \sum_{i=1}^{n} \{(1-p_i) I_{y_i>0}\} \\
&\quad + \sum_{i=1}^{n} \left\{ p_i \frac{(e^{-\lambda_i^+} - e^{-\lambda_i^-})}{(e^{-\lambda_i^-} + e^{(\alpha_0+\alpha_1 x_i)} e^{-\lambda_i^+})} I_{y_i=0} \right\} \\
\frac{\partial l}{\partial \alpha_1} &= \sum_{i=1}^{n} \{-p_i x_i I_{y_i<0}\} + \sum_{i=1}^{n} \{(1-p_i) x_i I_{y_i>0}\} \\
&\quad + \sum_{i=1}^{n} \left\{ p_i x_i \frac{(e^{-\lambda_i^+} - e^{-\lambda_i^-})}{(e^{-\lambda_i^-} + e^{(\alpha_0+\alpha_1 x_i)} e^{-\lambda_i^+})} I_{y_i=0} \right\} \\
\frac{\partial l}{\partial \beta_0^-} &= \sum_{i=1}^{n} \{(-y_i - \lambda_i^-) I_{y_i<0}\} - \sum_{i=1}^{n} \left\{ \frac{\lambda_i^- e^{-\lambda_i^-}}{(e^{-\lambda_i^-} + e^{(\alpha_0+\alpha_1 x_i)} e^{-\lambda_i^+})} I_{y_i=0} \right\} \\
\frac{\partial l}{\partial \beta_1^-} &= \sum_{i=1}^{n} \{(-y_i - \lambda_i^-) x_i I_{y_i<0}\} - \sum_{i=1}^{n} \left\{ \frac{\lambda_i^- e^{-\lambda_i^-} x_i}{(e^{-\lambda_i^-} + e^{(\alpha_0+\alpha_1 x_i)} e^{-\lambda_i^+})} I_{y_i=0} \right\} \\
\frac{\partial l}{\partial \beta_0^+} &= \sum_{i=1}^{n} \{(y_i - \lambda_i^+) I_{y_i>0}\} - \sum_{i=1}^{n} \left\{ \frac{\lambda_i^+ e^{(\alpha_0+\alpha_1 x_i)} e^{-\lambda_i^+}}{(e^{-\lambda_i^-} + e^{(\alpha_0+\alpha_1 x_i)} e^{-\lambda_i^+})} I_{y_i=0} \right\} \\
\frac{\partial l}{\partial \beta_1^+} &= \sum_{i=1}^{n} \{(y_i - \lambda_i^+) x_i I_{y_i>0}\} - \sum_{i=1}^{n} \left\{ \frac{\lambda_i^+ e^{(\alpha_0+\alpha_1 x_i)} e^{-\lambda_i^+} x_i}{(e^{-\lambda_i^-} + e^{(\alpha_0+\alpha_1 x_i)} e^{-\lambda_i^+})} I_{y_i=0} \right\}
\end{aligned}
$$

C.2 Second Derivatives

The second derivatives of the log–likelihood function (expression (7.4)) is needed for the observed information matrix of the 'Delta Method'. The second derivatives are given below.

$$
\frac{\partial^2 l}{\partial \alpha_0^2} = -\sum_{i=1}^{n} \{p_i(1-p_i)I_{y_i<0} + p_i(1-p_i)I_{y_i>0}\}
$$
$$
+ \sum_{i=1}^{n} \left\{ \frac{p_i(e^{-\lambda_i^+} - e^{-\lambda_i^-})[(1-p_i)e^{-\lambda_i^-} - p_i e^{(\alpha_0+\alpha_1 x_i)}e^{-\lambda_i^+}]}{(e^{-\lambda_i^-} + e^{(\alpha_0+\alpha_1 x_i)}e^{-\lambda_i^+})^2} \right\} I_{y_i=0}
$$

$$
\frac{\partial^2 l}{\partial \alpha_1^2} = -\sum_{i=1}^{n} x_i^2 \{p_i(1-p_i)I_{y_i<0} + p_i(1-p_i)I_{y_i>0}\}
$$
$$
+ \sum_{i=1}^{n} x_i^2 \left\{ \frac{p_i(e^{-\lambda_i^+} - e^{-\lambda_i^-})[(1-p_i)e^{-\lambda_i^-} - p_i e^{(\alpha_0+\alpha_1 x_i)}e^{-\lambda_i^+}]}{(e^{-\lambda_i^-} + e^{(\alpha_0+\alpha_1 x_i)}e^{-\lambda_i^+})^2} \right\} I_{y_i=0}
$$

$$
\frac{\partial^2 l}{\partial \alpha_0 \partial \alpha_1} = -\sum_{i=1}^{n} x_i \{p_i(1-p_i)I_{y_i<0} + p_i(1-p_i)I_{y_i>0}\}
$$
$$
+ \sum_{i=1}^{n} x_i \left\{ \frac{p_i(e^{-\lambda_i^+} - e^{-\lambda_i^-})[(1-p_i)e^{-\lambda_i^-} - p_i e^{(\alpha_0+\alpha_1 x_i)}e^{-\lambda_i^+}]}{(e^{-\lambda_i^-} + e^{(\alpha_0+\alpha_1 x_i)}e^{-\lambda_i^+})^2} \right\} I_{y_i=0}
$$

$$
\frac{\partial^2 l}{\partial \beta_0^{-2}} = -\sum_{i=1}^{n} \lambda_i^- \left\{ I_{y_i<0} + \frac{e^{-\lambda_i^-}[e^{-\lambda_i^-} + e^{(\alpha_0+\alpha_1 x_i)}e^{-\lambda_i^+}(1-\lambda_i^-)]}{(e^{-\lambda_i^-} + e^{(\alpha_0+\alpha_1 x_i)}e^{-\lambda_i^+})^2} I_{y_i=0} \right\}
$$

$$
\frac{\partial^2 l}{\partial \beta_1^{-2}} = -\sum_{i=1}^{n} \lambda_i^- x_i^2 \left\{ I_{y_i<0} + \frac{e^{-\lambda_i^-}[e^{-\lambda_i^-} + e^{(\alpha_0+\alpha_1 x_i)}e^{-\lambda_i^+}(1-\lambda_i^-)]}{(e^{-\lambda_i^-} + e^{(\alpha_0+\alpha_1 x_i)}e^{-\lambda_i^+})^2} I_{y_i=0} \right\}
$$

$$
\frac{\partial^2 l}{\partial \beta_1^- \partial \beta_0^-} = -\sum_{i=1}^{n} \lambda_i^- x_i \left\{ I_{y_i<0} + \frac{e^{-\lambda_i^-}[e^{-\lambda_i^-} + e^{(\alpha_0+\alpha_1 x_i)}e^{-\lambda_i^+}(1-\lambda_i^-)]}{(e^{-\lambda_i^-} + e^{(\alpha_0+\alpha_1 x_i)}e^{-\lambda_i^+})^2} I_{y_i=0} \right\}
$$

$$\frac{\partial^2 l}{\partial \beta_0^{+2}} = -\sum_{i=1}^n \left\{ \lambda_i^+ I_{y_i>0} \right\}$$
$$- \sum_{i=1}^n \left\{ \lambda_i^+ \frac{e^{(\alpha_0+\alpha_1 x_i)} e^{-\lambda_i^+} [e^{(\alpha_0+\alpha_1 x_i)} e^{-\lambda_i^+} + e^{-\lambda_i^-}(1-\lambda_i^+)]}{(e^{-\lambda_i^-} + e^{(\alpha_0+\alpha_1 x_i)} e^{-\lambda_i^+})^2} I_{y_i=0} \right\}$$

$$\frac{\partial^2 l}{\partial \beta_1^{+2}} = -\sum_{i=1}^n \left\{ \lambda_i^+ x_i^2 I_{y_i>0} \right\}$$
$$- \sum_{i=1}^n \left\{ \lambda_i^+ x_i^2 \frac{e^{(\alpha_0+\alpha_1 x_i)} e^{-\lambda_i^+} [e^{(\alpha_0+\alpha_1 x_i)} e^{-\lambda_i^+} + e^{-\lambda_i^-}(1-\lambda_i^+)]}{(e^{-\lambda_i^-} + e^{(\alpha_0+\alpha_1 x_i)} e^{-\lambda_i^+})^2} I_{y_i=0} \right\}$$

$$\frac{\partial^2 l}{\partial \beta_1^+ \partial \beta_0^+} = -\sum_{i=1}^n \left\{ \lambda_i^+ x_i I_{y_i>0} \right\}$$
$$- \sum_{i=1}^n \left\{ \lambda_i^+ x_i \frac{e^{(\alpha_0+\alpha_1 x_i)} e^{-\lambda_i^+} [e^{(\alpha_0+\alpha_1 x_i)} e^{-\lambda_i^+} + e^{-\lambda_i^-}(1-\lambda_i^+)]}{(e^{-\lambda_i^-} + e^{(\alpha_0+\alpha_1 x_i)} e^{-\lambda_i^+})^2} I_{y_i=0} \right\}$$

$$\frac{\partial^2 l}{\partial \beta_0^- \partial \alpha_0} = \sum_{i=1}^n \left\{ \frac{p_i \lambda_i^- e^{-\lambda_i^-} e^{-\lambda_i^+} [1 + e^{(\alpha_0+\alpha_1 x_i)}]}{(e^{-\lambda_i^-} + e^{(\alpha_0+\alpha_1 x_i)} e^{-\lambda_i^+})^2} I_{y_i=0} \right\}$$

$$\frac{\partial^2 l}{\partial \beta_0^- \partial \alpha_1} = \sum_{i=1}^n x_i \left\{ \frac{p_i \lambda_i^- e^{-\lambda_i^-} e^{-\lambda_i^+} [1 + e^{(\alpha_0+\alpha_1 x_i)}]}{(e^{-\lambda_i^-} + e^{(\alpha_0+\alpha_1 x_i)} e^{-\lambda_i^+})^2} I_{y_i=0} \right\}$$

$$\frac{\partial^2 l}{\partial \beta_1^- \partial \alpha_0} = \sum_{i=1}^n x_i \left\{ \frac{p_i \lambda_i^- e^{-\lambda_i^-} e^{-\lambda_i^+} [1 + e^{(\alpha_0+\alpha_1 x_i)}]}{(e^{-\lambda_i^-} + e^{(\alpha_0+\alpha_1 x_i)} e^{-\lambda_i^+})^2} I_{y_i=0} \right\}$$

$$\frac{\partial^2 l}{\partial \beta_1^- \partial \alpha_1} = \sum_{i=1}^n x_i^2 \left\{ \frac{p_i \lambda_i^- e^{-\lambda_i^-} e^{-\lambda_i^+} [1 + e^{(\alpha_0+\alpha_1 x_i)}]}{(e^{-\lambda_i^-} + e^{(\alpha_0+\alpha_1 x_i)} e^{-\lambda_i^+})^2} I_{y_i=0} \right\}$$

$$\frac{\partial^2 l}{\partial \beta_0^+ \partial \alpha_0} = -\sum_{i=1}^n \left\{ \frac{p_i \lambda_i^+ e^{-\lambda_i^-} e^{-\lambda_i^+} [1 + e^{(\alpha_0+\alpha_1 x_i)}]}{(e^{-\lambda_i^-} + e^{(\alpha_0+\alpha_1 x_i)} e^{-\lambda_i^+})^2} I_{y_i=0} \right\}$$

$$\frac{\partial^2 l}{\partial \beta_0^+ \partial \alpha_1} = -\sum_{i=1}^n x_i \left\{ \frac{p_i \lambda_i^+ e^{-\lambda_i^-} e^{-\lambda_i^+} [1 + e^{(\alpha_0+\alpha_1 x_i)}]}{(e^{-\lambda_i^-} + e^{(\alpha_0+\alpha_1 x_i)} e^{-\lambda_i^+})^2} I_{y_i=0} \right\}$$

$$\frac{\partial^2 l}{\partial \beta_1^+ \partial \alpha_0} = -\sum_{i=1}^n x_i \left\{ \frac{p_i \lambda_i^+ e^{-\lambda_i^-} e^{-\lambda_i^+} [1 + e^{(\alpha_0+\alpha_1 x_i)}]}{(e^{-\lambda_i^-} + e^{(\alpha_0+\alpha_1 x_i)} e^{-\lambda_i^+})^2} I_{y_i=0} \right\}$$

$$\frac{\partial^2 l}{\partial \beta_1^+ \partial \alpha_1} = -\sum_{i=1}^n x_i^2 \left\{ \frac{p_i \lambda_i^+ e^{-\lambda_i^-} e^{-\lambda_i^+} [1 + e^{(\alpha_0+\alpha_1 x_i)}]}{(e^{-\lambda_i^-} + e^{(\alpha_0+\alpha_1 x_i)} e^{-\lambda_i^+})^2} I_{y_i=0} \right\}$$

$$\frac{\partial^2 l}{\partial\beta_0^+\partial\beta_0^-} = -\sum_{i=1}^{n}\left\{\frac{p_i\lambda_i^-\lambda_i^+e^{-\lambda_i^-}e^{-\lambda_i^+}[1+e^{(\alpha_0+\alpha_1 x_i)}]}{(e^{-\lambda_i^-}+e^{(\alpha_0+\alpha_1 x_i)}e^{-\lambda_i^+})^2}I_{y_i=0}\right\}$$

$$\frac{\partial^2 l}{\partial\beta_0^+\partial\beta_1^-} = -\sum_{i=1}^{n}x_i\left\{\frac{p_i\lambda_i^-\lambda_i^+e^{-\lambda_i^-}e^{-\lambda_i^+}[1+e^{(\alpha_0+\alpha_1 x_i)}]}{(e^{-\lambda_i^-}+e^{(\alpha_0+\alpha_1 x_i)}e^{-\lambda_i^+})^2}I_{y_i=0}\right\}$$

$$\frac{\partial^2 l}{\partial\beta_1^+\partial\beta_0^-} = -\sum_{i=1}^{n}x_i\left\{\frac{p_i\lambda_i^-\lambda_i^+e^{-\lambda_i^-}e^{-\lambda_i^+}[1+e^{(\alpha_0+\alpha_1 x_i)}]}{(e^{-\lambda_i^-}+e^{(\alpha_0+\alpha_1 x_i)}e^{-\lambda_i^+})^2}I_{y_i=0}\right\}$$

$$\frac{\partial^2 l}{\partial\beta_1^+\partial\beta_1^-} = -\sum_{i=1}^{n}x_i^2\left\{\frac{p_i\lambda_i^-\lambda_i^+e^{-\lambda_i^-}e^{-\lambda_i^+}[1+e^{(\alpha_0+\alpha_1 x_i)}]}{(e^{-\lambda_i^-}+e^{(\alpha_0+\alpha_1 x_i)}e^{-\lambda_i^+})^2}I_{y_i=0}\right\}$$

C.3 Expected Values

Expected values of the second derivatives that were computed in the previous section are given below.

$$E\left(\frac{\partial^2 l}{\partial\alpha_0^2}\right) = -\sum_{i=1}^{n}p_i\left\{(1-p_i)-\frac{e^{-\lambda_i^-}e^{-\lambda_i^+}}{(e^{-\lambda_i^-}+e^{(\alpha_0+\alpha_1 x_i)}e^{-\lambda_i^+})}\right\}$$

$$E\left(\frac{\partial^2 l}{\partial\alpha_1^2}\right) = -\sum_{i=1}^{n}x_i^2 p_i\left\{(1-p_i)-\frac{e^{-\lambda_i^-}e^{-\lambda_i^+}}{(e^{-\lambda_i^-}+e^{(\alpha_0+\alpha_1 x_i)}e^{-\lambda_i^+})}\right\}$$

$$E\left(\frac{\partial^2 l}{\partial\alpha_1\partial\alpha_0}\right) = -\sum_{i=1}^{n}x_i p_i\left\{(1-p_i)-\frac{e^{-\lambda_i^-}e^{-\lambda_i^+}}{(e^{-\lambda_i^-}+e^{(\alpha_0+\alpha_1 x_i)}e^{-\lambda_i^+})}\right\}$$

$$E\left(\frac{\partial^2 l}{\partial\beta_0^{-2}}\right) = -\sum_{i=1}^{n}\lambda_i^-\left\{(1-p_i)-\frac{p_i\lambda_i^-e^{-\lambda_i^-}e^{-\lambda_i^+}}{(e^{-\lambda_i^-}+e^{(\alpha_0+\alpha_1 x_i)}e^{-\lambda_i^+})}\right\}$$

$$E\left(\frac{\partial^2 l}{\partial\beta_1^{-2}}\right) = -\sum_{i=1}^{n}x_i^2\lambda_i^-\left\{(1-p_i)-\frac{p_i\lambda_i^-e^{-\lambda_i^-}e^{-\lambda_i^+}}{(e^{-\lambda_i^-}+e^{(\alpha_0+\alpha_1 x_i)}e^{-\lambda_i^+})}\right\}$$

$$E\left(\frac{\partial^2 l}{\partial\beta_1^-\partial\beta_0^-}\right) = -\sum_{i=1}^{n}x_i\lambda_i^-\left\{(1-p_i)-\frac{p_i\lambda_i^-e^{-\lambda_i^-}e^{-\lambda_i^+}}{(e^{-\lambda_i^-}+e^{(\alpha_0+\alpha_1 x_i)}e^{-\lambda_i^+})}\right\}$$

$$E\left(\frac{\partial^2 l}{\partial \beta_0^{+2}}\right) = -\sum_{i=1}^{n} p_i \lambda_i^+ \left\{ 1 - \frac{\lambda_i^+ e^{-\lambda_i^-} e^{-\lambda_i^+}}{\left(e^{-\lambda_i^-} + e^{(\alpha_0 + \alpha_1 x_i)} e^{-\lambda_i^+}\right)} \right\}$$

$$E\left(\frac{\partial^2 l}{\partial \beta_1^{+2}}\right) = -\sum_{i=1}^{n} x_i^2 p_i \lambda_i^+ \left\{ 1 - \frac{\lambda_i^+ e^{-\lambda_i^-} e^{-\lambda_i^+}}{\left(e^{-\lambda_i^-} + e^{(\alpha_0 + \alpha_1 x_i)} e^{-\lambda_i^+}\right)} \right\}$$

$$E\left(\frac{\partial^2 l}{\partial \beta_1^+ \partial \beta_0^+}\right) = -\sum_{i=1}^{n} x_i p_i \lambda_i^+ \left\{ 1 - \frac{\lambda_i^+ e^{-\lambda_i^-} e^{-\lambda_i^+}}{\left(e^{-\lambda_i^-} + e^{(\alpha_0 + \alpha_1 x_i)} e^{-\lambda_i^+}\right)} \right\}$$

$$E\left(\frac{\partial^2 l}{\partial \beta_0^- \partial \alpha_0}\right) = \sum_{i=1}^{n} \frac{p_i \lambda_i^- e^{-\lambda_i^-} e^{-\lambda_i^+}}{\left(e^{-\lambda_i^-} + e^{(\alpha_0 + \alpha_1 x_i)} e^{-\lambda_i^+}\right)}$$

$$E\left(\frac{\partial^2 l}{\partial \beta_0^- \partial \alpha_1}\right) = \sum_{i=1}^{n} x_i \frac{p_i \lambda_i^- e^{-\lambda_i^-} e^{-\lambda_i^+}}{\left(e^{-\lambda_i^-} + e^{(\alpha_0 + \alpha_1 x_i)} e^{-\lambda_i^+}\right)}$$

$$E\left(\frac{\partial^2 l}{\partial \beta_1^- \partial \alpha_0}\right) = \sum_{i=1}^{n} x_i \frac{p_i \lambda_i^- e^{-\lambda_i^-} e^{-\lambda_i^+}}{\left(e^{-\lambda_i^-} + e^{(\alpha_0 + \alpha_1 x_i)} e^{-\lambda_i^+}\right)}$$

$$E\left(\frac{\partial^2 l}{\partial \beta_1^- \partial \alpha_1}\right) = \sum_{i=1}^{n} x_i^2 \frac{p_i \lambda_i^- e^{-\lambda_i^-} e^{-\lambda_i^+}}{\left(e^{-\lambda_i^-} + e^{(\alpha_0 + \alpha_1 x_i)} e^{-\lambda_i^+}\right)}$$

$$E\left(\frac{\partial^2 l}{\partial \beta_0^+ \partial \alpha_0}\right) = -\sum_{i=1}^{n} \frac{p_i \lambda_i^+ e^{-\lambda_i^-} e^{-\lambda_i^+}}{\left(e^{-\lambda_i^-} + e^{(\alpha_0 + \alpha_1 x_i)} e^{-\lambda_i^+}\right)}$$

$$E\left(\frac{\partial^2 l}{\partial \beta_0^+ \partial \alpha_1}\right) = -\sum_{i=1}^{n} x_i \frac{p_i \lambda_i^+ e^{-\lambda_i^-} e^{-\lambda_i^+}}{\left(e^{-\lambda_i^-} + e^{(\alpha_0 + \alpha_1 x_i)} e^{-\lambda_i^+}\right)}$$

$$E\left(\frac{\partial^2 l}{\partial \beta_1^+ \partial \alpha_0}\right) = -\sum_{i=1}^{n} x_i \frac{p_i \lambda_i^+ e^{-\lambda_i^-} e^{-\lambda_i^+}}{\left(e^{-\lambda_i^-} + e^{(\alpha_0 + \alpha_1 x_i)} e^{-\lambda_i^+}\right)}$$

$$E\left(\frac{\partial^2 l}{\partial \beta_1^+ \partial \alpha_1}\right) = -\sum_{i=1}^{n} x_i^2 \frac{p_i \lambda_i^+ e^{-\lambda_i^-} e^{-\lambda_i^+}}{\left(e^{-\lambda_i^-} + e^{(\alpha_0 + \alpha_1 x_i)} e^{-\lambda_i^+}\right)}$$

$$E\left(\frac{\partial^2 l}{\partial \beta_0^- \partial \beta_0^+}\right) = -\sum_{i=1}^{n} \frac{p_i \lambda_i^- \lambda_i^+ e^{-\lambda_i^-} e^{-\lambda_i^+}}{\left(e^{-\lambda_i^-} + e^{(\alpha_0 + \alpha_1 x_i)} e^{-\lambda_i^+}\right)}$$

$$E\left(\frac{\partial^2 l}{\partial \beta_0^- \partial \beta_1^+}\right) = -\sum_{i=1}^{n} x_i \frac{p_i \lambda_i^- \lambda_i^+ e^{-\lambda_i^-} e^{-\lambda_i^+}}{\left(e^{-\lambda_i^-} + e^{(\alpha_0 + \alpha_1 x_i)} e^{-\lambda_i^+}\right)}$$

$$E\left(\frac{\partial^2 l}{\partial \beta_1^- \partial \beta_0^+}\right) = -\sum_{i=1}^{n} x_i \frac{p_i \lambda_i^- \lambda_i^+ e^{-\lambda_i^-} e^{-\lambda_i^+}}{\left(e^{-\lambda_i^-} + e^{(\alpha_0 + \alpha_1 x_i)} e^{-\lambda_i^+}\right)}$$

$$E\left(\frac{\partial^2 l}{\partial \beta_1^- \partial \beta_1^+}\right) = -\sum_{i=1}^{n} x_i^2 \frac{p_i \lambda_i^- \lambda_i^+ e^{-\lambda_i^-} e^{-\lambda_i^+}}{\left(e^{-\lambda_i^-} + e^{(\alpha_0 + \alpha_1 x_i)} e^{-\lambda_i^+}\right)}$$

CURRICULUM VITAE

Rasitha Rangnai Jayasekare Kodippuli Thanthillage Dona
Department of Mathematics
University of Louisville, Louisville, KY 40292
email : rkjaya01@louisville.edu
phone : (502) - 852 - 7012

Education:

- B.Sc. in Applied Science, 2004, Rajarata University of Sri Lanka, Sri Lanka

- M.Sc. in Industrial Mathematics, 2008, University of Sri Jayewardenepura, Sri Lanka
  Thesis "Optimal Utilization of Machines in an Apparel Production Line"

- M.A. in Mathematics, 2011, University of Louisville, USA.

Teaching Experience:

- Graduate Teaching Assistant : Department of Mathematics, University of Louisville, USA. (August 2009 - July 2013)

- Lecturer : School of Computing, Asia Pacific Institute of Information Technology, Sri Lanka, (January 2006 - July 2009)

- Lecturer : Department of Physical Sciences, Faculty of Applied Sciences, Rajarata University of Sri Lanka, (September 2004 - December 2005)

Papers:

- K.T.D.R.R. Wickramasinghe*, D.D.A. Gamini, B.M.S.G Banneheka, Optimal Utilization of Machines in an Apparel Production Line. Published for the 50th Anniversary Academic Conference, University of Sri Jayewardenepura Sri Lanka, 2009.

Achievements:

- Recognized at the Deans Reception for participating graduate professional and career workshops in 2013.

- Awarded as one of the Top Two Graduate Talks at the 26th Annual Eastern Kentucky University Symposium in the Mathematical, Statistical and Computer Sciences at Eastern Kentucky University, Kentucky, in March 2013.

- Ken F. and Sandra S. Hohman Fellowship for Excellent Class Work and Diligent Teaching of Department of Mathematics at University of Louisville for years 2011 2012.

- Gold Medal for the Best Performance in the Department of the Department of Physical Sciences, Rajarata University of Sri Lanka, 2004.

Presentations:

- Poster presentation on Detecting Significant Changes in Stock Price using a Liquidity Effect Model at the Graduate Research Symposium of University of Louisville, Kentucky, in February 2012.

- A presentation on Multiple Change Point Estimation in a Liquidity Effect Model at the Mathematics Association of America (MAA) Kentucky section meeting at Bellarmine University, Kentucky, in March 2012.

- A presentation on Application of Finite Mixture Model involving Poisson Distribution at the 32nd Annual Mathematics Symposium at Western Kentucky University, Kentucky, in October 2012.

- A presentation on Understanding Changes in Stock Price Using a Finite Mixture at the 26th Annual Eastern Kentucky University Symposium in the Mathematical, Statistical and Computer Sciences at Eastern Kentucky University, Kentucky, in March 2013.

- A presentation on Poisson Mixture Model for Discrete Stock Price Changes at the Mathematics Association of America (MAA) Kentucky section meeting at Transylvania University, Kentucky, in April 2013.

Competitions:

- Participated in the group data mining project to provide a solution to Influenza Impact for SAS Data Mining Shootout Competition 2012. SAS Enterprise Guide 4.3 and Enterprise Miner 7.1 were used.

Certifications:

- Certificate in Information Technology at the completion of the first year of Bachelor of Information Technology (BIT) at University of Colombo, Sri Lanka in 2003.

- Advance Certificate in Information Technology at the completion of the second year of Bachelor of Information Technology (BIT) at University of Colombo, Sri Lanka in 2004.