Two Essays on Actuarial and Financial Econometrics

by
Kwadwo Asamoah

A dissertation submitted in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

in

Economics

Stony Brook University

December, 2015

Stony Brook University

The Graduate School

Kwadwo Asamoah

We, the dissertation committee for the above candidate for the Doctor of Philosophy degree, hereby recommend acceptance of this dissertation.

Mark R. Montgomery - Dissertation Advisor
Professor, Department of Economics
Stony Brook University

Samuele Centorrino - Chairperson of Defense
Assistant Professor, Department of Economics
Stony Brook University

Jake Zhao - Committee Member
Assistant Professor, Department of Economics
Stony Brook University

Aaron Kim- Outside Committee Member
Assistant Professor, College of Business,
Stony Brook University

This dissertation is accepted by the Graduate School

Charles Taber

Dean of the Graduate School

Abstract of the Dissertation

Two Essays on Actuarial and Financial Econometrics

by

Kwadwo Asamoah

Doctor of Philosophy

in

Economics

Stony Brook University

2015

In chapter 1, we analyze the concept of credibility in two generalized count models: Mittag leffler and Weibull count models which can handle both underdispersion and overdispersion in count data and nest the commonly used Poisson model as a special case. A correct specification of the model is important since without a proper pricing mechanism one is simply not competitive in the insurance industry. We find evidence using data from Danish Insurance Company that the simple Poisson model can set the credibility weight to one resulting from large heterogeneity among polycyholders and thus breaks the credibility model down. We propose parametric estimators for the structural parameters in the credibility formula using the mean and variance of the assumed distributions and a maximum likelihood estimation over a collective data. As an example, we show that the proposed parameters from Mittag leffler provides weights that are consistent with the idea of credibility whiles a simulation study is carried out to investigate the stability of the maximum likelihood estimates from the Weibull count model. Finally, we extend analyses to multidimensional lines and show how our approach can be adopted to cross selling application of the credibility model.

Chapter 2 shows that Multinomial Logit (MNL) previously found to be a good tool in predicting performance in the Indian market can only be implemented at the industry level but not the entire U.S. market if the appropriate financial ratios are selected with low predictive power but consistent with the efficiency of the market. We design a multilayer perceptron (MLP) neural network and show that as an alternative tool for the U.S. market when prediction is of ultimate importance with overall average accuracy rate at about 57.6% and 59.4% in a training and testing data respectively. The results obtained reveal that a firm's ability to pay its short term obligations and how efficient it uses cash for generating sales revenue are highly predictive and important when predicting stock performance in a probabilistic framework.

Dedicated to my mum and family

TABLE OF CONTENTS

List of Figures

List of Tables

LIST OF APPENDICES

Appendix

LIST OF ABBREVIATIONS

ADRs American Depository Receipts

ANNs Artificial Neural Networks

B/M book-per-market value

BVPS book value per share

CAPM capital asset pricing model

CFS cash flow per share

CT cash turnover

CURRENT current ratio

DMKT debt to market ratio

DY dividend yield

EBIT earnings-before-interest-and-tax

EMH Efficient Market Hypothesis

E/P earnings-price ratio

EPS earnings per share

FLR financial leverage ratio

GMM Generalized Method of Moments

MDA Multivariate Discriminant Analysis

MLEs maximum likelihood estimates

MLP multilayer perceptron

MNL Multinomial Logit

MLR multinomial logistic regression

NPM net profit margin

NYSE New York Stock Exchange

OLS Ordinary Least Squares

P/D price-dividend ratio

P/E price-earnings ratio

P/B price-book ratio

ROA return on assets

ROI return on investments

Acknowledgements

I am extremely grateful to my advisor Zari Rachev for his tremendous guidance and encouragement through out my research work. I would also like to thank Mark R. Montgomery, Thomas Muench, Aaron Kim and Samuele Centorrino for their continuous help and support. Finally, a lot of appreciation goes out to Jens P. Nielsen, Martin Englund and Blake McShane for all their support.

CHAPTER I

On the Credibility of Insurance Claim Frequency: Generalized Count Models and Parametric Estimators

1.1 Introduction

The basic idea underlying insurance is that individuals transfer their risks to insurance companies and in return pay periodic premium. The premium paid by the insured is usually a combination of a risk premium, also called pure risk, and risk loadings, which include an extra amount for other expenses, profit, and a margin for contingencies. Risk premiums are calculated by estimating individual's risk levels, which are expectations of some loss measures. These measures could be the number of claims made by a customer per policy year (claim frequency), claim severity or the total claim amount.

In this study, we focus on claim frequency information as a predictor of individuals' risk levels which may depend on current period information as well as past experience. The determination of a fair risk premium is a major concern in the actuarial profession. One classical approach in calculating the individual premium is to use the expectation principle (Pinquet, 1997, 1998; Desjardins et al., 2001). In this approach, the correct individual premium in a given year will be the expected value of his claim frequency.[1]

An alternative approach, which we follow in this research, is the linear credibility theory. The claim experience observed for an individual policyholder is usually too limited to be statistically reliable in predicting the pure risk. However, every individual risk is usually part of a risk class so that the collective claim history in a large class can provide credible statistical predictions.[2] Therefore, it is clear that both sources of information should be used in calculating the fair risk premium.

Credibility assigns meaningful weights to the individual and collective experience estimates of the pure risk premium. When there is enough information on the individual claims history, the credibility weight on the individual estimate increases. The credibility concept dates back to Mowbray (1914) and Whitney (1918) and has since been given attention in the literature. There have been several studies on the extensions of this model usually in a non-parametric setting including Bühlmann (1967), Jewell (1973, 1974), Hachemeister (1975), Sundt (1979, 1981) and Zehnwirth (1985). This paper is focused on the approach presented by Bühlmann and Straub (1970), which eliminated the limitation in the Bühlmann model by allowing the process variance of the loss measure (claim frequency in this case) to depend on the exposure.[3]

The number of claims, given the individual unobserved risk profile, is generally assumed to follow a Poisson distribution (Bichsel, 1967; Pinquet, 1997; Desjardins et al., 2001; Bühlmann and Gisler, 2005; Englund et al., 2008).[4] However, a Poisson distribution is valid only under a very restrictive assumption of equidispersion in the data. A Poisson model implies that given an unobserved risk profile, the number of claims made by an individual policyholder has no variability (the mean equals the

[1]There are other specifications of the expected value principle to include the risk loadings such as the standard deviation, variance and exponential principles.

[2]Usually risks are classified into homogeneous classes (risk classes) based on some quantifiable characteristics of policyholders such as age, occupation, sex, geographical location among others.

[3]Exposure as defined by American Institute for Chartered Property Casualty Underwriters is any condition that presents a possibility of loss, regardless of whether loss actually occurs.

[4]In fact, every study on the applications of credibility theory known to us has modeled claim frequencies with this so-called standard assumption.

variance). This assumption may or may not hold true for every policyholder. In practice, a company may deal with millions of customers and may be over ambiguous to assume that the number of claims for each one is Poisson distributed.[5] A misspecification of the distribution of the loss measure could lead to mispricing and therefore unfair premium paid by policyholders.

In addition, count data are usually the outcomes of an underlying count process in continuous time. A Poisson distribution is derived under the assumption that the number of events in a given time period results from exponentially distributed inter-arrival times, implying a constant hazard function. In practice, the occurrence of an event may increase the probability of occurrence of the next event (positive dependence) or reduce the possibility of the next event (negative dependence). Seal (1969) and Pinquet (2000) show that positive dependence in claim counts characterizes the automobile insurance industry.

Statisticians, having recognized this limitation of the Poisson distribution, have developed numerous count models with the inter-arrival times following Gamma, Weibull, Mittag Leffler and many other distributions, thereby allowing for overdispersion and underdispersion in the data. A heterogeneous Gamma-Poisson model (negative binomial) was first presented in Greenwood and Yule (1920) and it is routinely used to model overdispersed count data. A number of other studies including King (1989), Winkelman (1995), Cameron and Johansson (1997) and Cameron and Trivedi (1998) have since developed different models which address the issue of underdispersion.

In a recent study, McShane et al. (2008) have derived a generalized count data model based upon a Weibull inter-arrival time process that nests the Poisson and negative binomial models as special cases.[6] The Weibull count model, via the shape

[5]As we report later in the paper, we see all kinds of claim counts pattern in the data used for this study.

[6]This new model is usually referred to as Weibull count model.

parameter being less than, equal to, or greater than 1, can handle overdispersed, equidispersed, and underdispersed data respectively and is computationally tractable. A similar generalized count model has been developed by Kanichukattu and Abraham (2011) when the inter-arrival times follow a Mittag Leffler distribution, referred to as Mittag Leffler count model. The Mittag Leffler count model also nests the Poisson count model and has the ability to handle overdispersed and underdispersed data.

This paper introduces the Weibull and Mittag Leffler count models in a Bühlmann-Straub credibility model and studies the effects of the Poisson restriction. In addition to the non-parametric estimation methods for the structural parameters in the literature, this work investigates and presents alternative parametric estimators using the mean and variance of the assumed count distributions and maximum likelihood estimation procedures. This paper contributes to the literature by giving flexibility to the credibility formula and its parameters to handle many kinds of data processes. We begin the analysis in the one dimensional model and then extend it to the multi-dimensional case and explains how it can be adopted to the cross-selling application of the model.

The remainder of the paper is organized as follows. Section 1.2 presents the standard Bühlmann-Straub credibility model in a Weibull count model and shows the corresponding parametric estimators with a simulation study. Section 1.3 undertakes a similar analyses in a Mittag Leffler count model. A multidimensional version of the models in the previous two sections is provided in Section 1.4. In Section 1.5, we present the empirical findings and the data used. Finally, Section 1.6 concludes.

1.2 Bühlmann-Straub credibility model in Weibul Count Model

1.2.1 One Bühlmann-Straub credibility model

We analyze the classical one dimensional Bühlmann-Straub credibility model for claim frequency when the inter-arrival time of the occurrence of claims N_{ij} for customer i in year j on one policy follow the Weibull distribution.[7]

Let N_{ij} be the number of claims observed for customer i in year j on one policy. We make the standard assumption in the literature that, we have a time dependent covariates which give estimated prior information on the expected number of claims of the customers each period. The estimated prior knowledge λ_{ij} of customer i in period j is modeled as

$$\lambda_{ij} = w_{ij} g(X'_{ij}) \tag{1.1}$$

Where X_{ij} is the vector of prior explanatory variables of a policyholder i in year j and w_{ij} is the number of years at risk (duration).

We again assume that each customer i has his own risk profile θ_i, which is a realization of the random variable Θ_i, then following from the Bühlmann-Straub credibility model assumptions we have that:

Assumption I.1.

1. *$N_{ij}|\Theta_i$ are independent and follow the Weibull count model for $j = 1, \cdots J$ with*

$$E(N_{ij}|\Theta_i) = \lambda_{ij}\mu(\Theta_i) \tag{1.2}$$

$$Var(N_{ij}|\Theta_i) = \lambda_{ij}\sigma^2(\Theta_i) \tag{1.3}$$

Where $\mu(\Theta_i)$ and $\sigma^2(\Theta_i)$ are the mean and variance of the Weibull count model.

[7] We will use the same notations as in the original model and in the literature so that we can make reference to results therein.

2. *The pairs* $(\Theta_1, \boldsymbol{N}_1)$, $(\Theta_2, \boldsymbol{N}_2)$,... *are independent and* Θ_i *(*$i = 1, 2, ...$*) are iid with* $\Theta_i \sim gamma(r, \alpha)$.

Given $N_{i1}, N_{i2}, ..., N_{iJ}$, we want to find the best linear predictor for N_{ij+1}. The estimator function is chosen to be the posterior mean equal to

$$E(\mu(\theta)|N_{i1}, N_{i2}, ..., N_{iJ})$$

If we define the claim frequency F_{ij} of customer i in year j to be $\frac{N_{ij}}{\lambda_{ij}}$, then F_{ij} satisfy the hypothesis of the Bühlmann-Straub model and therefore given $N_{i1}, ...N_{iJ}$, the best linear credibility estimator of customer i in year $J+1$ is

$$\begin{aligned} \widehat{\mu(\Theta_i)} &= E(\mu(\Theta_i)) + [Var(\mu(\Theta_i))]^2 \left(\frac{E(\sigma^2(\Theta_i))}{\lambda_i Var(\mu(\Theta_i))} + 1 \right)^{-1} (F_i - E(F_i)) \\ &= \mu_0 + \eta_i (F_i - \mu_0) \end{aligned} \tag{1.4}$$

where

$$\eta_i = \frac{\lambda_i}{\lambda_i + \kappa}, \quad \kappa = \frac{\sigma^2}{\tau^2}, \quad \mu_0 = E(\mu(\Theta_i)), \quad \sigma^2 = E(\sigma^2(\Theta_i)), \quad \tau^2 = Var(\mu(\Theta_i)),$$
$$\lambda_i = \sum_{j=1}^{J} \lambda_{ij} \text{ and } F_i = \sum_{j=1}^{J} \frac{\lambda_{ij}}{\lambda_i} F_{ij}.$$

The above result is straightforward from minimizing the mean square error. Detailed derivation of this result is repeated in Appendix A.

1.2.2 Parameters Estimation

We need to estimate the three structural parameters μ_0, σ^2 and τ^2 in order to obtain the credibility estimator for each customer.[8] The consequence of the Poisson distribution assumption of the number of claims in the literature is that the portfolio premium μ_0 is the same as the average variance within the individual risk σ^2.

[8] μ_0, σ^2 and τ^2 are respectively the portfolio premium, the average variance within individual customer's risk and the variance between individual risks.

There are several studies in the actuarial literature on the best estimates for these parameters (Dubey and Gisler, 1981; Norberg, 1982).[9] In this work, we provide an alternative to the estimation suggestions in the literature by using maximum likelihood and the mean and variance of the Weibull count model (see Appendix B for the moments of the Weibull count model). By our assumption of the Weibull count model, we have that the probability of observing n from an individual with a risk profile Θ_i is

$$P(N = n|\Theta_i) = \sum_{m=n}^{\infty} \frac{(-1)^{m+n}\alpha_m^n \Theta_i^m}{\Gamma(cm+1)} \tag{1.5}$$

For $n = 0, 1, 2, ...$ Where $\alpha_m^0 = \frac{\Gamma(cm+1)}{\Gamma(m+1)}$, for $m = 0, 1, 2, ...$ and $\alpha_m^{n+1} = \sum_{l=n}^{m-1} \frac{\Gamma(cm-cl+1)}{\Gamma(m-l+1)}$, for $n = 0, 1, 2, ...$ for $m = n+1, n+2, n+3, ...$

The risk profile of a policyholder is assumed to be homogeneous in time and as a result, the structural parameters are time-independent. Englund et al. (2009) extend this model to include time effects and report no significant change in the estimators. This is because observed claim histories are usually available for only a short period of time, usually under five years and policyholders' risks do not change. In a pure Bayesian analysis, these parameters are mostly estimated based on the opinions of experts. They can also be determined from collective observations of similar risks. [10]

The individual unobserved risk profiles are independent and identically distributed random variables drawn from the gamma distribution (i.e. $\Theta_i \sim gamma(r, \alpha)$). If we integrate across all possible risk profiles in any given year, we have the heterogeneous

[9] Usually the estimation methods suggested are non-parametric and are used irrespective of the observed statistical information (number of claims, claim sizes or claim amount). However, when there is a structure in the model, it should be considered. A typical example is when the number of claims is Poisson-distributed, we need to consider the fact that the estimate for μ_0 should be theoretically the same as that for σ^2. Bühlmann and Gisler (2005) incorporates this structure by some iterative procedure.

[10] This approach is referred to as the empirical Bayes. Since the estimators are time independent, we can use collective information from any year.

Weibull count model (McShane et al., 2008) given by

$$P(N = n) = \sum_{m=n}^{\infty} \frac{(-1)^{m+n}\alpha_m^n}{\Gamma(cm+1)} \frac{\Gamma(r+m)}{\Gamma(r)\alpha^m} \tag{1.6}$$

For $n = 0, 1, 2, ...$ Where $\alpha_m^0 = \frac{\Gamma(cm+1)}{\Gamma(m+1)}$, for $m = 0, 1, 2, ...$ and $\alpha_m^{n+1} = \sum_{l=n}^{m-1} \frac{\Gamma(cm-cl+1)}{\Gamma(m-l+1)}$, for $n = 0, 1, 2, ...$ for $m = n+1, n+2, n+3, ...$ (The derivation of the heterogeneous model is repeated here in Appendix C). The heterogeneous count model describes the distribution over the collective data and we will refer to it as the structural or collective function.

Now, we have a considerably larger information for any year based on the collective observed number of claims n_i, $i = 1, ..., I$ to estimate the parameters c, r and α by maximum likelihood. The log likelihood function for $n = (n_1, \cdots, n_I)$ is

$$\begin{aligned} L(c, r, \alpha) &= \sum_{i=1}^{I} \ln P(N = n_i) \\ &= \sum_{i=1}^{I} \ln \sum_{m=n_i}^{\infty} \frac{(-1)^{m+n_i}\alpha_m^{n_i}}{\Gamma(cm+1)} \frac{\Gamma(r+m)}{\Gamma(r)\alpha^m} \end{aligned} \tag{1.7}$$

Let $\hat{c}$, $\hat{r}$ and $\hat{\alpha}$ be the maximum likelihood estimators. We will then estimate the structural parameters using the mean and variance form from the Weibull count model. The estimates for μ_0, σ^2 and τ^2 are respectively given by equations (1.8), (1.9) and (1.10) below:

$$\hat{\mu}_0 = \sum_{n=1}^{\infty} \sum_{m=n}^{\infty} \frac{n(-1)^{m+n}\alpha_m^n}{\Gamma(\hat{c}m+1)} \frac{\Gamma(\hat{r}+m)}{\Gamma(\hat{r})\hat{\alpha}^m} \tag{1.8}$$

$$\hat{\sigma}^2 = \sum_{n=1}^{\infty} \sum_{m=n}^{\infty} \frac{n^2(-1)^{m+n}\alpha_m^n}{\Gamma(\hat{c}m+1)} \frac{\Gamma(\hat{r}+m)}{\Gamma(\hat{r})\hat{\alpha}^m} - \widehat{E}\left(\sum_{n=1}^{\infty} \sum_{m=n}^{\infty} \frac{n(-1)^{m+n}\alpha_m^n}{\Gamma(\hat{c}m+1)} \Theta_i^m\right)^2 \tag{1.9}$$

$$\hat{\tau}^2 = \widehat{Var}\left(\sum_{n=1}^{\infty}\sum_{m=n}^{\infty}\frac{n(-1)^{m+n}\alpha_m^n}{\Gamma(\hat{c}m+1)}\Theta_i^m\right) \tag{1.10}$$

A detailed derivation of these estimators is provided in Appendix C. The expectation and variance in (1.9) and (1.10) can be estimated using Monte Carlo estimation method. With the maximum likelihood estimates $\hat{r}$ and $\hat{\alpha}$, we know the estimated gamma distribution of Θ_i as $gamma(r, \alpha)$. We will therefore simulate n samples of Θ_i and use Monte Carlo as demonstrated in Appendix D.

1.2.3 Simulation Study

The approach adapted for estimating the structural parameters in this paper depends highly on the maximum likelihood estimates (MLEs): $\hat{r}$, $\hat{\alpha}$ and $\hat{c}$ from the Weibull heterogeneity model. Even though the MLEs have good asymptotic properties and perform well in finite samples, their computation and implementation can be problematic in high dimensional situations. Therefore, we carry out a simulation study to determine the sensitivity of the estimates to different sample sizes and parameter values. As pointed out by McShane et al. (2008), the log likelihood function is robust to truncation points; 50 and 100.[11]

We undertake the bootstrapping in the following steps:

1. Guess some true parameter values for r, α and c and simulate I number of observations from the collective function

2. Build the log likelihood function and optimize to obtain the MLEs of the parameters

3. Repeat steps 1 and 2 n times with the same parameter values so that we have a sample distribution of the estimates.

4. Finally, find the mean, quantiles and obtain a box plot for the distributions

[11] We have verified this claim and can confirm that it is true.

First, we consider $r = 5$, $\alpha = 4$ and $c = 2$ as our true parameters. Using these parameters, we assume four(4) different sample sizes (3500, 5000, 8000, 10000) and simulate observations from the collective function. The mean and median of samples distribution of the estimates presented in Table E.1 show that the maximum likelihood estimation is doing well and the results are consistent across the different samples. This means we can estimate the parameters well irrespective of whether we have large or small sample size. Figures E.1–E.4 representing the box plots of the sample distributions clearly show a fairly normal distribution.

We then change the true parameters to $r = 7$, $\alpha = 3$ and $c = 1$ and do the same experiment for the different sample sizes. The results again suggest that with the right starting values, the maximum likelihood estimation does a good job in estimating the true parameters despite the complexity of the log likelihood function. We see that the estimates are pretty close to the true parameters as shown in Table E.2 and Figures E.5–E.8

Finally, we consider another set of true parameters: $r = 7$, $\alpha = 3$ and $c = 0.95$ and repeat previous exercise. Notice that since $c = 0.95$ the resulting samples display overdispersion. Like the previous samples, the results for this scenario as shown in in Table E.3 and Figures E.9–E.12 also show that the maximum likelihood method estimates the true parameters well. The set of parameters are chosen to underdispersion, equidispersion and overdispersion data and in all cases we see that the estimates are closer to the true parameters

1.3 Bühlmann-Straub credibility model in Mittag Leffler Count Model

1.3.1 One Bühlmann-Straub credibility model

Now, we analyze the one dimensional Bühlmann-Straub credibility model for claim frequency when the interarrival time of the occurrence of claims N_{ij} for customer i in year j on one policy follow the Mittag Leffler distribution. The resulting count model is Mittag leffler count model (Kanichukattu and Abraham, 2011).

The problem, we want to solve here is the same as Section 1.2.1 except that now the number of claims of customer i given its risk profile is assumed to follow the Mittag Leffler count model. Therefore, the Bühlmann-Straub credibility model assumptions become:

Assumption I.2.

1. *$N_{ij}|\Theta_i$ are independent and follow the Mittag Leffler count model for $j = 1, \cdots J$ with*

$$E(N_{ij}|\Theta_i) = \lambda_{ij}\mu(\Theta_i) \tag{1.11}$$

$$Var(N_{ij}|\Theta_i) = \lambda_{ij}\sigma^2(\Theta_i) \tag{1.12}$$

   Where $\mu(\Theta_i)$ and $\sigma^2(\Theta_i)$ are the mean and variance of the Mittag Leffler count model.

2. *The pairs $(\Theta_1, \boldsymbol{N}_1)$, $(\Theta_2, \boldsymbol{N}_2)$,... are independent and Θ_i $(i = 1, 2, ...)$ are iid with $\Theta_i \sim gamma(r, \alpha)$.*

1.3.2 Parameters Estimation

Again, one needs to be able to estimate the three structural parameters μ_0, σ^2 and τ^2. This can be done using the already existing non-parametric estimators. However, we show how parametric estimators can be derived. The process here is similar to what was done in the case when the number of claims were assumed to follow the Weibull count model. With our assumption, the probability of observing n from an individual with a risk profile Θ_i is

$$P(N = n|\Theta_i) = \sum_{m=n}^{\infty} \frac{(-1)^{m-n}\binom{m}{n}\Theta_i^{cm}}{\Gamma(cm+1)} \tag{1.13}$$

Now, we can obtain the collective function $P(N = n)$ by integrating across all the possible risk profiles in a given year. From the collective function below we can build the log likelihood function using data from the entire portfolio and estimate c, r and α using maximum likelihood. See Appendix G for the derivation of the Mittag Leffler collective function.

$$P(N = n) = \sum_{m=n}^{\infty} \frac{(-1)^{m-n}\binom{m}{n}}{\Gamma(r)\alpha^{cm}} \frac{\Gamma(cm+r)}{\Gamma(cm+1)} \tag{1.14}$$

For $n = 0, 1, 2, ...$ The log likelihood function to estimate r, α and c using the collective function from Mittag Leffler for $n = (n_1, \cdots, n_I)$ is

$$\begin{aligned} L(c, r, \alpha) &= \sum_{i=1}^{I} \ln P(N = n_i) \\ &= \sum_{i=1}^{I} \ln \sum_{m=n_i}^{\infty} \frac{(-1)^{m-n}\binom{m}{n}}{\Gamma(r)\alpha^{cm}} \frac{\Gamma(cm+r)}{\Gamma(cm+1)} \end{aligned} \tag{1.15}$$

After getting the MLEs, $\hat{c}$, $\hat{r}$ and $\hat{\alpha}$ and using the mean and variance of the Mittag Leffler count model (see Appendix F), we obtain the parametric estimators for the structural parameters as follows:

$$\widehat{\mu_0} = \frac{\Gamma(\hat{r}+\hat{c})}{\Gamma(1+\hat{c})\Gamma(\hat{r})\hat{\alpha}^{\hat{c}}} \tag{1.16}$$

$$\widehat{\sigma^2} = \frac{1}{\Gamma(\hat{r})\hat{\alpha}^{\hat{c}}}\left(\frac{\Gamma(\hat{r}+\hat{c})}{\Gamma(1+\hat{c})} + \frac{2\Gamma(\hat{r}+2\hat{c})}{\Gamma(1+2\hat{c})\hat{\alpha}^{\hat{c}}} - \frac{\Gamma(\hat{r}+2\hat{c})}{\Gamma(1+\hat{c})^2\hat{\alpha}^{\hat{c}}}\right) \tag{1.17}$$

$$\widehat{\tau^2} = \frac{1}{\Gamma(1+\hat{c})^2\Gamma(\hat{r})\hat{\alpha}^{2\hat{c}}}\left(\Gamma(\hat{r}+2\hat{c}) - \frac{\Gamma(\hat{r}+\hat{c})^2}{\Gamma(\hat{r})}\right) \tag{1.18}$$

1.4 Multidimensional Bühlmann-Straub

In this section, we extend the model presented in Sections 1.2.1 and 1.3.1 by increasing the dimensions of the observed number of claims. There are many situations in the actuarial industry where it becomes necessary to consider several business lines for a particular customer or different types of claims for a policy held by a customer. The latter could be normal claims versus big claims (Bühlmann et al., 2003). Normally, there are few biggest claims which account for more than half of the total claim amount and hence they are sometimes treated separately from the normal claims.

The second case considers different business lines for one customer (Englund et al., 2008).[12] In this paper, we follow Bühlmann and Gisler (2005) and Englund et al. (2008) and assumes one specific risk variable for each product.

Suppose each customer buys L products with an insurance company in the multivariate environment. Let $\theta_i = (\theta_{i1}, ..., \theta_{iL})$ be a vector of risk profiles of customer i corresponding to the different policies held by him. These unobserved risks are realizations of a random vector $\Theta_i = (\Theta_{i1}, ..., \Theta_{iL})$. The estimated prior differences now have three subscripts and thus ρ_{ijl} represents the information of customer i in the year j for the product type l.

The credibility estimator in this setting is obtained similarly to the univariate case by understanding the projection operator componentwise. It is straightforward

[12] This is particularly useful because depending on the correlation between the different risk profiles one can use information from one business line to price products in another where such information is not available. One risk profile for all the different business lines implies a correlation of 1. However, in practice one could observe the existence of different types of positive correlation.

to derive the best linear credibility formula as:

$$\widehat{\boldsymbol{\mu}(\Theta_i)} = E(\boldsymbol{\mu}(\Theta_i)) + Cov(\boldsymbol{\mu}(\Theta_i), \boldsymbol{\mu}(\Theta_i)')Cov(\mathbf{F}_i, \mathbf{F}_i')^{-1}(\mathbf{F}_i - E(\mathbf{F}_i)) \tag{1.19}$$

Where now $\widehat{\boldsymbol{\mu}(\Theta_i)} = \left(\widehat{\mu(\Theta_{i1})}, ..., \widehat{\mu(\Theta_{iL})}\right)'$ with the components being the linear credibility estimators for the different lines of business and $\mathbf{F}_i$ is an L-dimensional vector such that the lth component equals $\sum_{j=1}^{J} \frac{\rho_{ijl}}{\rho_{il}} F_{ijl}$ and $\rho_{il} = \sum_{j=1}^{J} \rho_{ijl}$

The multidimensional credibility estimator in equation (3.1) can be simplified as a weighted sum of individual and collective claim histories as:

$$\widehat{\boldsymbol{\mu}(\Theta_i)} = (\mathbf{I} - \boldsymbol{\eta}_i)\boldsymbol{\mu}_0 + \boldsymbol{\eta}_i \mathbf{F}_i \tag{1.20}$$

Where

$\boldsymbol{\mu}_0 = E(\boldsymbol{\mu}(\Theta_i))$, $\boldsymbol{\eta} = \mathbf{T}(\mathbf{T} + \mathbf{S}\boldsymbol{\rho}_i^{-1})^{-1}$, $\mathbf{T} = Cov(\boldsymbol{\mu}(\Theta_i), \mathbf{F}_i') = Cov(\boldsymbol{\mu}(\Theta_i), \boldsymbol{\mu}(\Theta_i)')$ and $\mathbf{S} = E[Cov(\mathbf{F}_i, \mathbf{F}_i'|)\Theta_i]$.

$\boldsymbol{\rho}_i$ is a $L \times L$ diagonal matrix with the lth element $\rho_{il} = \sum_{j=1}^{J} \rho_{ijl}$ and $\mathbf{I}$ is the $L \times L$ identity matrix.

1.4.1 Parameters Estimation

1.4.1.1 Weibul Count Model

The parameters to be estimated in the multivariate case are $\boldsymbol{\mu}_0$, $\mathbf{S}$ and $\mathbf{T}$. Our approach seeks to provide alternative parametric estimators to the distribution-free estimators in the literature (DeVylder, 1978; Bühlmann and Gisler, 2005). We will estimate the parameters element by element using the same idea as in the one dimensional case. Let μ_{0l} be the collective premium corresponding to the lth business line. Then, we can find the maximum likelihood estimates $\hat{c}_l$, $\hat{r}_l$ and $\hat{\alpha}_l$ by using the observations across the different policyholders for product l so that

$$\mu_{0l} = \sum_{n=1}^{\infty} \sum_{m=n}^{\infty} \frac{n(-1)^{m+n} \alpha_m^n}{\Gamma(\hat{c}_l m + 1)} \frac{\Gamma(\hat{r}_l + m)}{\Gamma(\hat{r}_l) \hat{\alpha}_l^m} \tag{1.21}$$

For $l = 1, ..., L$. Putting these elements together, we get the collective premium vector

$$\boldsymbol{\mu}_0 = (\mu_{01}, ..., \mu_{0L})' \tag{1.22}$$

We have assumed that, given the unobserved risk profile parameter Θ_i, the claim frequencies $\mathbf{F}_i$ are independent. Therefore $\mathbf{S}$ is an $L \times L$ diagonal matrix and the lth diagonal element is given by

$$\sigma_l^2(\Theta_i) = \sum_{n=1}^{\infty} \sum_{m=n}^{\infty} \frac{n^2(-1)^{m+n} \alpha_m^n}{\Gamma(cm + 1)} \Theta_{il}^m - \left(\sum_{n=1}^{\infty} \sum_{m=n}^{\infty} \frac{n(-1)^{m+n} \alpha_m^n}{\Gamma(cm + 1)} \Theta_{il}^m \right)^2 \tag{1.23}$$

By using the information from product l we can get the maximum likelihood estimates and estimate these elements $\sigma_l^2(\Theta_i)$, for $l = 1, ..., L$ by:

$$\hat{\sigma}_l^2 = \sum_{n=1}^{\infty} \sum_{m=n}^{\infty} \frac{n^2(-1)^{m+n} \alpha_m^n}{\Gamma(\hat{c}_l m + 1)} \frac{\Gamma(\hat{r}_l + m)}{\Gamma(\hat{r}_l) \hat{\alpha}_l^m} - \widehat{E} \left(\sum_{n=1}^{\infty} \sum_{m=n}^{\infty} \frac{n(-1)^{m+n} \alpha_m^n}{\Gamma(\hat{c}_l m + 1)} \Theta_{il}^m \right)^2 \tag{1.24}$$

where the expectation in the estimators is approximated using Monte Carlo (see Appendix D).

Next, we need to estimate the covariance matrix $\mathbf{T}$. The diagonal elements can be estimated in the same way as in the one dimensional model. An estimate for the lth diagonal element of $\mathbf{T}$ is

$$\hat{\tau}_l^2 = \widehat{Var} \left(\sum_{n=1}^{\infty} \sum_{m=n}^{\infty} \frac{n(-1)^{m+n} \alpha_m^n}{\Gamma(\hat{c}_l m + 1)} \Theta_{il}^m \right) \tag{1.25}$$

Where $\hat{c}_l$, $\hat{r}_l$ and $\hat{\alpha}_l$ are estimated by maximum likelihood using information for the

lth business line. The non-diagonal elements of $\mathbf{T}$ for $l \neq l'$ is given by

$$\begin{aligned} cov(\mu(\Theta_{il}), \mu(\Theta_{il'})) &= E(\mu(\Theta_{il}).\mu(\Theta_{il'})) - E(\mu(\Theta_{il}))E(\mu(\Theta_{il'})) \\ &= E(\mu(\Theta_{il}).\mu(\Theta_{il'})) - \mu_{0l}\mu_{0l'} \end{aligned} \tag{1.26}$$

The non-diagonal elements are particularly important for instance, in a two dimensional model where an employer insures his employees both 'at work' and 'not at work', the non-diagonal elements cannot be zero since it is known that people who are risky at work, tend to be more risky outside work. With the MLEs, the non-diagonal elements can be estimated as:

$$\widehat{cov}(\mu(\Theta_{il}), \mu(\Theta_{il'})) = \widehat{E}(\mu(\Theta_{il}).\mu(\Theta_{il'})) - \hat{\mu}_{0l}\hat{\mu}_{0l'} \tag{1.27}$$

where the expectation can be estimated with Monte Carlo.

1.4.1.2 Mittag Leffler Count Model

When the Mittag Leffler count model is assumed, we follow the same routine exercise to provide parametric estimators for $\boldsymbol{\mu}_0$, $\mathbf{S}$ and $\mathbf{T}$.

The lth component of $\boldsymbol{\mu}_0$ is given by

$$\widehat{\mu}_{0l} = \frac{\Gamma(\hat{r}_l + \hat{c}_l)}{\Gamma(1+\hat{c}_l)\Gamma(\hat{r}_l)\hat{\alpha}_l^{\hat{c}_l}} \tag{1.28}$$

Again, $\mathbf{S}$ is diagonal with the lth element estimated using data on the lth product given by

$$\widehat{\sigma}_l^2 = \frac{1}{\Gamma(\hat{r}_l)\hat{\alpha}_l^{\hat{c}_l}} \left(\frac{\Gamma(\hat{r}_l + \hat{c}_l)}{\Gamma(1+\hat{c}_l)} + \frac{2\Gamma(\hat{r}_l + 2\hat{c}_l)}{\Gamma(1+2\hat{c}_l)\hat{\alpha}_l^{\hat{c}_l}} - \frac{\Gamma(\hat{r}_l + 2\hat{c}_l)}{\Gamma(1+\hat{c}_l)^2\hat{\alpha}_l^{\hat{c}_l}} \right) \tag{1.29}$$

Next, we need to estimate the between variance matrix $\mathbf{T}$. The diagonal elements are straightforward treating each element in the same way as the one dimensional estimation procedure. Following from the one dimensional procedure, the lth diagonal element of $\mathbf{T}$ is

$$\widehat{\tau}_l^2 = \frac{1}{\Gamma(1+\hat{c}_l)^2\Gamma(\hat{r}_l)\hat{\alpha}_l^{2\hat{c}_l}}\left(\Gamma(\hat{r}_l+2\hat{c}_l)-\frac{\Gamma(\hat{r}_l+\hat{c}_l)^2}{\Gamma(\hat{r}_l)}\right) \tag{1.30}$$

The non-diagonal elements use information from two different products. Consider any two products l and l' such that $l \neq l'$ then the corresponding element is

$$\begin{aligned} cov(\mu(\Theta_{il}),\mu(\Theta_{il'})) &= E(\mu(\Theta_{il}).\mu(\Theta_{il'})) - E(\mu(\Theta_{il}))E(\mu(\Theta_{il'})) \\ &= E\left(\frac{\Theta_{il}^{c_l}}{\Gamma(1+c_l)}.\frac{\Theta_{il'}^{c_{l'}}}{\Gamma(1+c_{l'})}\right) - \mu_{0l}\mu_{0l'} \\ &= \frac{1}{\Gamma(1+c_l)\Gamma(1+c_{l'})}E\left(\Theta_{il}^{c_l}\Theta_{il'}^{c_{l'}}\right) - \mu_{0l}\mu_{0l'} \end{aligned} \tag{1.31}$$

which can be estimated as:

$$\widehat{cov}(\mu(\Theta_{il}),\mu(\Theta_{il'})) = \frac{1}{\Gamma(1+\hat{c}_l)\Gamma(1+\hat{c}_{l'})}\widehat{E}\left(\Theta_{il}^{\hat{c}_l}\Theta_{il'}^{\hat{c}_{l'}}\right) - \hat{\mu}_{0l}\hat{\mu}_{0l'} \tag{1.32}$$

1.4.2 Application to Cross-Selling

The multidimensional credibility model studied here has been recently applied to cross selling by Kaishev et al. (2013), Thuring et al. (2012) and Thuring (2010). Cross-selling is simply approaching the present customers of a company and encouraging them to purchase one or more additional products of the company. The idea is to identify less risky customers based on their history with the company and sell them new products.

Cross-selling is important for a company to lower customers' churn rate, increasing the number of loyal customers and obtaining higher customer lifetime value (Akura and Srinivasan, 2005) and the different product features allow significant contributions for managers striving for valuable and strong relationship with their current customer base (Larivière and den Poel, 2004). It also helps companies to learn more about the customers' preferences and buying behavior, accumulate various types of data to their database and use such information as a predictor of certain behaviors of the customers such as customer retention and profitability outcomes (Ahn et al., 2011; Larivière and den Poel, 2005; Kamakura et al., 2003).

Previous studies assume equidispersion at individual claims level and consequently assume that the number of claims follow a Poisson distribution, which means selecting customers is based on the risk profile θ_{il} for customer i with product l, which is also the predicted number of claims. They also use the distribution free estimators by Englund et al. (2008) and Bühlmann and Gisler (2005). With these assumptions for example, Thuring et al. (2012) failed to identify 20% of the customers to target. This is because after estimating the risk profiles, ordering them in ascending order, the deviation between the estimated priori expected number of claims and the observed number of claims was the lowest for the group with least estimated risk profiles.

The approach presented in this work gives so much flexibility in this area. First, the restricted Poisson assumption can be generalized and customers can be selected based on the means of more generalized distributions which are functions of the risk profiles. Consider the multidimensional credibility formula obtained Section 1.4, the estimator for the lth product can written as:

$$\begin{aligned}\mu(\Theta_{il}) &= \mu_{0l} + \sum_{l'=1}^{L} \eta_{ill'}(F_{il'} - \mu_{0l'}) \\ &= \mu_{0l} + \eta_{ill}(F_{il} - \mu_{0l}) + \sum_{l' \neq l} \eta_{ill'}(F_{il'} - \mu_{0l'})\end{aligned} \tag{1.33}$$

In addition to the distribution flexibility, the parametric estimators provided in Section 1.4.1 can also be exploited and it is highly recommended.

1.5 Empirical Studies

1.5.1 Data

The data for testing the methodology and hypothesis in this paper is taken from a large Danish Insurance company, consisting of individuals who have been customers of the company from 1999 to 2004. We have information on a total of 99,951 unique customers in the database. Th insurance information for each customer in the data include the expected (estimated) claim frequency, number of reported claims, duration (risk exposure), accident year, purchase date of policy, expiry date of policy and 11 different products or policies including building, traffic risks, motor hull and car. Table 1.1 below shows the distribution of the customers across the different years. We

Table 1.1: Distribution of Customers

Year	Number of Customers
1999	62,767
2000	59,544
2001	61,500
2002	60,248
2003	60,867
2004	60,102

will randomly divide the data set of each product into two samples: estimation data and validation data.The estimation data will consist of 75% of the total customers after removing NAs.[13] The remaining 25% will constitute the validation data.

1.5.2 Results: Generalized Vrs. Poisson Count Models

The results presented here assess the distributional assumptions on the non-parametric estimators by Bühlmann and Gisler (2005) using data from personal building and chattels lines of business. It is important to point out not all policyholders

[13] The NAs in the data indicate that the particular policies are not owned by the customers.

have owned the lines of business for the same the number of years. We therefore, select customers who have held the policies for the entire period under consideration.

We first consider customers who have held the personal building and chattels coverages for the entire period from 1999 to 2004. The final samples consist of 2823 and 12681 policyholders for personal chattels and building coverages respectively. Table 1.2 compares the structural parameters under two scenarios: when we assume a generalized count model (Weibull and Mittag-leffler count models) and when we assume that individual personal chattels claim counts follow the Poison model whiles Table 1.3 shows the results for personal building. For details of the non-parametric estimators including the iterative procedure imposed by the Poisson assumption, refer to Bühlmann and Gisler (2005).

Table 1.2: Structural Parameters: Personal Chattels (1999-2004)

	Bühlmann-Gisler	
Parameters	Generalized Count Models	Poisson model
μ_0	0.656	0.591
$\sigma^2 = E(\sigma^2(\Theta))$	0.503	0.591
$\tau^2 = Var(\mu(\Theta))$	3.802	1564300

Table 1.3: Structural Parameters: Personal Building (1999-2004)

	Bühlmann-Gisler	
Parameters	Generalized Count Models	Poisson Model
μ_0	0.881	0.819
$\sigma^2 = E(\sigma^2(\Theta))$	0.875	0.819
$\tau^2 = Var(\mu(\Theta))$	0.560	60682496

Assuming equidispersion at individual claims level has a serious effect on the between variance parameter in the credibility formula. When the number of claims of a policyholder is assumed to be Poison, the restriction imposed by equating the collective premium (μ_0) and the within variance parameter (σ^2), simply renders the credibility formula unimportant. The Poisson model gives full credence to the individual observed number of claims even for 6 years. However, when the equidispersion assumption is relaxed, we get credibility weights that suggests that 6 years claims information on the individual customer is not enough and accordingly assigns higher weights to the collective information. Evidence using sample from the personal chattels validation data sample is provided in Table 1.4. While the Poisson model gives

Table 1.4: Credibility Weights: Personal Chattels (1999-2004)

	Credibility Weight	
	Generalized Count Models	Poisson model
$N_{ij} = \{0, 0, 0, 0, 0, 0\}$ $\lambda_{ij} = \{0.003, 0.003, 0.003, 0.003, 0.003, 0.002\}$	0.104	1.00
$N_{ij} = \{1, 0, 0, 0, 0, 0\}$ $\lambda_{ij} = \{0.010, 0.009, 0.002, 0.010, 0.008, 0.011\}$	0.274	1.00
$N_{ij} = \{0, 1, 0, 0, 0, 0\}$ $\lambda_{ij} = \{0.003, 0.003, 0.003, 0.011, 0.001, 0.011\}$	0.199	1.00
$N_{ij} = \{0, 0, 1, 0, 0, 0\}$ $\lambda_{ij} = \{0.005, 0.005, 0.005, 0.005, 0.003, 0.005\}$	0.176	1.00
$N_{ij} = \{0, 0, 0, 1, 0, 0\}$ $\lambda_{ij} = \{0.002, 0.018, 0.020, 0.020, 0.003, 0.020\}$	0.381	1.00
$N_{ij} = \{0, 0, 0, 0, 1, 0\}$ $\lambda_{ij} = \{0.007, 0.012, 0.012, 0.009, 0.014, 0.014\}$	0.332	1.00
$N_{ij} = \{0, 0, 0, 0, 0, 1\}$ $\lambda_{ij} = \{0.004, 0.007, 0.007, 0.007, 0.007, 0.007\}$	0.231	1.00
$N_{ij} = \{0, 1, 1, 0, 0, 0\}$ $\lambda_{ij} = \{0.012, 0.012, 0.009, 0.011, 0.011, 0.011\}$	0.330	1.00

100% credibility to individual observed claim experience, the highest credibility weight assigned to the individual experience without the Poisson restriction is 38.1%. The lowest weight is assigned to the case where the individual reported no claim for the entire six year period.

Now, suppose we have only three years of claims experience about the customers from 2002 to 2004. Table 1.5 below compares the structural parameters under the different models. Since the experience information has decreased, the estimated structural parameters responded accordingly. However, the between variance, $\hat{\tau}^2$ resulting from the heterogeneity among customers is still large enough to break the credibility model down as shown in Table 1.6.

Table 1.5: Structural Parameters: Personal Chattels (2002-2004)

	Bühlmann-Gisler	
Parameters	Generalized Count Models	Poisson model
μ_0	0.6035106	0.4360533
$\sigma^2 = E(\sigma^2(\Theta))$	0.4305286	0.4360533
$\tau^2 = Var(\mu(\Theta))$	0.8617676	851632.2

We see that using the generalized count models, the Bühlmann and Gisler (2005) estimators methods are able to adjust and reduce the weights assigned to the individual experience while the Poisson model still gives full credence to 3 years experience.

Table 1.6: Credibility Weights: Personal Chattels (2002-2004)

	Credibility Weight	
	Generalized Count Models	Poisson model
$N_{ij} = \{0,0,0,0,0,0\}$ $\lambda_{ij} = \{0.003, 0.003, 0.003, 0.003, 0.003, 0.002\}$	0.015	1.00
$N_{ij} = \{1,0,0,0,0,0\}$ $\lambda_{ij} = \{0.010, 0.009, 0.002, 0.010, 0.008, 0.011\}$	0.055	1.00
$N_{ij} = \{0,1,0,0,0,0\}$ $\lambda_{ij} = \{0.003, 0.003, 0.003, 0.011, 0.001, 0.011\}$	0.044	1.00
$N_{ij} = \{0,0,1,0,0,0\}$ $\lambda_{ij} = \{0.005, 0.005, 0.005, 0.005, 0.003, 0.005\}$	0.024	1.00
$N_{ij} = \{0,0,0,1,0,0\}$ $\lambda_{ij} = \{0.002, 0.018, 0.020, 0.020, 0.003, 0.020\}$	0.077	1.00
$N_{ij} = \{0,0,0,0,1,0\}$ $\lambda_{ij} = \{0.007, 0.012, 0.012, 0.009, 0.014, 0.014\}$	0.067	1.00
$N_{ij} = \{0,0,0,0,0,1\}$ $\lambda_{ij} = \{0.004, 0.007, 0.007, 0.007, 0.007, 0.007\}$	0.041	1.00
$N_{ij} = \{0,1,1,0,0,0\}$ $\lambda_{ij} = \{0.012, 0.012, 0.009, 0.011, 0.011, 0.011\}$	0.062	1.00

1.5.3 Results: Mittag Leffler Parametric Vrs. Bühlmann-Gisler Structural Parameters

We want to estimate the parametric estimators from the Mittag Leffler count model and compare that to the non-parametric estimators. One can think of this estimation method as a 2-stage process. In stage 1, we find MLEs of the parameters of the distribution of Θ to be used as the prior distribution in the second stage using the most recent collective information (2004) of all customers and the collective function (1.14). The MLEs from the personal chattels estimation sample are $\hat{r} = 0.010$, $\hat{\alpha} = 10.233$ and $\hat{c} = 0.587$.

Since $\hat{c} = 0.587$, the collective sample is over-dispersed and it can easily be confirmed by comparing the mean and variance. With these estimates, we obtain the parametric estimators from equations (1.16)– (1.18) in the 2 stage. The results are provided in the last column of Table 1.7. The collective premium obtained is relatively small, however this is compensated by a corresponding higher weight compared to the Bühlmann-Gisler estimators when the individual equidispersion assumption is generalized. The credibility weights on a validation data is presented in Table 1.8.

The Mittag Leffler estimators suggest that 6 years experience data on a customer is not credible enough and accordingly assigns higher weights to the collective information. A nice property of these estimators is that they share similar qualitative properties as the generalized Bühlmann-Gisler estimators. For examples, both estimators rank the samples selected in the same increasing order of their credibility weights.

Table 1.7: Structural Parameters (with Mittag Leffler): Personal Chattels (1999-2004)

Parameters	Bühlmann-Gisler		
	Generalized Count	Poisson model	Mittag-Leffler
μ_0	0.656	0.591	0.004
$\sigma^2 = E(\sigma^2(\Theta))$	0.503	0.591	0.005
$\tau^2 = Var(\mu(\Theta))$	3.802	1564300	0.001

Table 1.8: Credibility Weights (with Mittag Leffler): Personal Chattels (1999-2004)

	Credibility Weight		
	Generalized Count Models	Mittag Leffler	Poisson model
$N_{ij} = \{0, 0, 0, 0, 0, 0\}$	0.104	0.003	1.00
$N_{ij} = \{1, 0, 0, 0, 0, 0\}$	0.274	0.010	1.00
$N_{ij} = \{0, 1, 0, 0, 0, 0\}$	0.199	0.007	1.00
$N_{ij} = \{0, 0, 1, 0, 0, 0\}$	0.176	0.006	1.00
$N_{ij} = \{0, 0, 0, 1, 0, 0\}$	0.381	0.016	1.00
$N_{ij} = \{0, 0, 0, 0, 1, 0\}$	0.332	0.013	1.00
$N_{ij} = \{0, 0, 0, 0, 0, 1\}$	0.231	0.008	1.00
$N_{ij} = \{0, 1, 1, 0, 0, 0\}$	0.330	0.013	1.00

1.6 Conclusion

This work demonstrates that the usual restricted Poisson assumption of modeling clam frequencies in credibility theory can be eliminated. Under some assumptions we show that the nice properties of two generalized count models (the Weibull and Mittag Leffler count models) for handling all kinds of data in terms variability can be carried to the credibility estimator of individual risk levels.

In addition to existing estimation methods for the structural parameters in the credibility formula, we have contributed to the literature by providing an alternative parametric estimators using the mean and variance of the assumed distribution of the number of claims given individual's risk profile and a maximum likelihood over a collective data. The Mittag Leffler structural parameters provide credibility weights that have similar qualitative empirical properties as those provided by Bühlmann and

Gisler (2005) without the Poisson restriction.

we further show that these analyses can be extended to the multidimensional case and we show how elements in the structural parameters can be estimated. It is important to mention that the maximum likelihood approach provided in this work relies heavily on the time invariant assumption of the unobserved risk profile which has also been shown to be the case in the literature, unless there is enough information on the individual policyholder.

In this paper, we draw the attention of actuaries to the distribution assumptions that are made about the number of claims used in determining individual risk premiums. The premium can be determined fairly when the right distribution is used. We introduce the Weibull count model and the Mittag Leffler count model in credibility which handles many kinds of data processes. We conclude by saying that a thorough investigation into what distribution fits a company claim data process is very important since as evidenced in this work, the Poisson model though easy to work with, can lead to misleading results.

CHAPTER II

Predicting Performance for Stocks Selection in U.S. Markets

2.1 Introduction

The U.S. markets are considered to be 'efficient', an idea commonly known as the Efficient Market Hypothesis (EMH) which asserts that stocks always trade at their fair value on financial exchanges, making it impossible for investors to either sell inflated or purchase undervalued stocks and thus the only way to possibly earn higher returns is by purchasing riskier assets. While the EMH is often disputed by many, academics point to a large body of evidence in support of it. Past studies have suggested financial ratios as an important tool for predicting stock performance which is reflected in the stock's return. Company's annual report which are made available through accounting principles provide enough financial data which transformed into various ratios. These ratios have proved to be valuable not only in determining a company's relative performance but also are important tool in forecasting future performance.

Investors take risks each time they enter the securities market. The decision of which stock to buy vary among investors depending on their objectives.[1] Some

[1] These may be short term or long term. Investors with short term goals may be interested in companies paying frequent and high dividends.

depend on public information to decide which securities they should buy; others use sophisticated models which they hope will give them an upper edge in such a competitive market like the stock exchange. Some people succeed whiles many more loose everything despite these models.[2]

The performance rates of the different stocks in ones portfolio may be different with some performing more poorly compared to market. In fact, predicting stock performance is very complicated and difficult but may be impossible. This paper employs key financial ratios in fundamental analysis to predict the performance of U.S. listed stocks in a probabilistic framework. The tool developed in this work will be useful for fund managers, investment companies as well as individual investors in making their investment decisions.

Financial investments play a significant role in the U.S. economy. Movements in the market can have serious effects on investors and hence the economy. The stock markets serve as a major source of employment for many whiles most people depend on the securities markets for their retirements. For some households, the returns from financial securities may be their sole source of income. It is therefore important for shareholders and other major stakeholders to use relevant financial information to enable them to invest in good securities in the stock market.[3]

The remainder of this paper is organized into the six sections. Section 2.2 presents a brief literature review. The third section 2.3 expands upon the relevant economic theory and its relation to the performance measure. Section 2.4 presents the empirical models and the estimation techniques employed. The fourth section 2.5 will be devoted to describing the data set used for the analysis, while Section 2.6 presents the empirical results and interpretations. The last section 2.7 contains conclusion and proposes suggestions for future research.

---

[2]'About two-thirds of all active investors will under perform index funds every year' Taylor (2004).

[3]Stakeholders basically include potential investors, employees, customers, suppliers and government.

2.2 Literature Review

Since the beginning of financial trading in the 12th century, several studies have attempted to predict the market in order for investors to gain competitive advantage over one another. Financial ratios were mostly used by banks and other lending institutions to assess a company's ability to pay short-term debt. Additionally, they were used as a benchmark for a competitive business analysis. However, financial analysts have realized the importance of these ratios in making investment decisions and predicting future performance of an entity.

Analysts may use financial ratios for forecasting future return trends and can give an early warning of a firm's deteriorating financial condition (Ohlson, 1980). For instance, financially hard-pressed companies can be correctly identified with a 94% accuracy rate within two years prior to the declaration of bankruptcy by assessing financial ratios (Altman, 1968).[4] Over the past years, several empirical studies have demonstrated the importance of these ratios contrary to the suggestion that the ratio analysis is no longer an important analytical tool in academia (Altman, 1968; Turk, 2006).

Most of the studies have focused on identifying which ratios affect expected stock returns or changes in stock prices. Cochrane (1997) combined Ordinary Least Squares (OLS) and Generalized Method of Moments (GMM) in his study and found that price-dividend ratio (P/D) can predict long run stock returns.[5] In a similar study, Lewellen (2004) showed that dividend yield (DY), book-per-market value (B/M) and earnings-price ratio (E/P) are good indications of a company's future stock price in current economic environment.

Hobarth (2006) studied the relationship between financial indicators and firm's

[4] He only considered publicly held manufacturing companies for which financial data were available. He outlined this situation as a limitation to his paper.

[5] He got decent results even though he did not include key ratios like price-earnings ratio (P/E) and price-book ratio (P/B). Also, his findings are true for longer time horizons over 5-10 years.

performance of listed firms in US using OLS. The result shows that companies with low B/M, efficient working capital, more equity and less debt, negative stock rating, few assets, high earnings-before-interest-and-tax (EBIT) margin and high profitability will have a better market performance measured by stock price (Hobarth, 2006). Other authors have shown the impact of several ratios on company's stock return but the strongest of the indicators in U.S. markets have been P/E, DY and P/B . However, most of these studies have focused on "point forecast" and a large portion of expected returns are left unexplained.[6]

The idea of using financial data in the probabilistic framework has been well researched in the literature but authors have mainly concentrated on predicting a company's failure or bankruptcy rate. Altman (1968) developed a default-prediction model using Multivariate Discriminant Analysis (MDA) and introduced the zscore rate of bankruptcy where higher values greater than 3.0 indicate low probability of bankruptcy with lower values suggesting otherwise. Ohlson (1980) criticized the MDA model, particularly the restrictive statistical requirements on the distributional properties of the explanatory variables and revisited the problem using logistic regression. Zavgren (1985) built a similar logit model that gives a five-year prior indication to a company's failure and found profitability and turnover ratios to be a significant indicator.

Following up on the concerns about the MDA model (Altman, 1968) and the logit model (Ohlson, 1980), others techniques including probit, recursive partitioning, hazard models and neural networks have been used in the prediction of a company's failure rate (Zmijewski, 1984; Jones, 1987; Agarwal and Taffler, 2007). Surprisingly, the question of how these ratios affect the probability of a company performing good or bad has been given little or no attention particularly in the U.S. markets. Upadhyay

[6]This issue has been recognized by the Vanguard Research Group and states: ... But the fact that even P/Es-the strongest of the indicators we examined-leave a large portion of returns unexplained underscores our belief that expected stock returns are best stated in a probabilistic framework, not as a "point forecast" (Joseph Davis and Thomas, 2012)

et al. (2012) pioneered the probabilistic prediction of a company's performance for the Indian Stock Market. In a MNL model, they predicted 56.8% of the sample correct and achieved a 58.41% accuracy prediction rate for out of sample data using seven (7) financial ratios, for the years from 2005 to 2008.[7]

In their limited study, only 30 companies with large market capitalization and are part of the Nifty index were selected resulting in a relatively small sample size of 118 and 101 for training and testing respectively. It is however not true that if a model works for top established companies then it must work for the entire market and thus, the restrictive nature of the selection we believe could lead to a possible sample selection bias. Another concern of their study which the authors admit is the issue of multicollinearity which makes it impossible to assess and interpret the effects of the ratios on the performance probabilities. However, they stated that their objective was to show that the MNL model could be used in predicting performance and a such the coefficients themselves didn't matter to them as long as they predicted the outcomes correctly.

This paper builds on the study by Upadhyay et al. (2012) and extends the analysis to a more efficient market. First, we attempt to answer the question of how one could use the MNL model in predicting performance for stocks listed on U.S. markets. Since one cannot interpret the results when multicollinearity is present, and so if predictability is preferred, It will be interesting to compare the MNL model with a neural network model. In addition, the frequency of the data could play a role and so one of the goals of the research work is to compare the results of annual data to that of a quarterly data.

In addition, we will bring the analysis to the industry level and attempt to understand how MNL performs. Financial ratios may have different impacts depending

[7]They identified percentage increase in net sales, earnings per share, book value, price/earnings per share, profit before interest depreciation and taxes/sales, price/book value and percentage change in operating profit as the key ratios for prediction in the Indian market.

on the economic sector of the company in question. For instance, a shoe store will have goods that quickly lose value because of changing fashion trends. However, these goods are easily sold and have high turnover. As a result, small amounts of money continuously come in and go out, and in a worst-case scenario liquidation is relatively simple. This company could easily function with a current ratio close to 1.0. An airplane manufacturer on the other hand, should have much higher current ratio to allow for coverage of short-term liabilities. Also, some financial ratios may be more predictive in some industries and thus in this study, we would want to identify some important ratios to look out for when investing in the energy, industrials and informational technology sectors.

One distinguishing feature of this research which is important to point out is how performance is measured. Unlike Upadhyay et al. (2012), we use a performance measure which is appropriate when selecting stocks to form a portfolio. To address the sample selection issue, every listed company is selected and to account for their differences, we control for their market capitalization in the model. Finally, this work will propose a set of financial ratios one should consider when predicting performance in a probabilistic framework for the U.S. markets.

2.3 Theory and Performance Measure

2.3.1 Theoretical Framework

Conceptually, the performance of a stock is measured by the returns it gives to investors. The rate of return from holding a stock equals the sum of capital gains (the change in price) plus any cash dividend payments, divided by the initial purchase price of the security.[8] Mathematically we have that,

$$R_t = \frac{P_{t+1} - P_t + D_t}{P_t} \tag{2.1}$$

Where:

R_t is the rate of return on the stock held from time t to t+1

P_{t+1} is the price of the stock at time t+1

P_t is the price of the stock at time t

D_t is the dividend payments made in the period t to t+1

Under the EMH. The capital asset pricing model (CAPM) gives a fair expected risk premium on a stock as:

$$r_s - r_f = \beta(r_m - r_f) \tag{2.2}$$

Where:

r_s = expected return on stock

r_f = risk free rate

r_m = expected market return

β = beta of the stock[9]

Upadhyay et al. (2012) developed performance by comparing a stock's return and variance to that of the market. While that works for the Indian market, we think the story is different when considering the U.S. market. Figures H.1 and H.2 compare the distributions of stock returns between the two markets.[10] Clearly, the skewness of expected returns in the U.S. market does suggest that we cannot use the variance alone as risk in differentiating among the performance groups. The expected returns

[8] Different investors may have different goals for buying stocks. People who depend on their investments for daily income would be more interested in cash dividends.

[9] β is the co-variance between the stock and the market returns divided by the variance of the market return. An asset has $\beta = 0$ if it is uncorrelated with the market. It has positive β if it follows the benchmark(market) and if β is negative, the return on the asset moves in opposite direction to the market return.

[10] The stock returns for the Indian stock market consist of the 30 companies as in Upadhyay et al. (2012) from year 2005 to 2008. We also consider the same sample period for NYSE.

selected for the Indian market are at least close to a normal distribution. In fact, Shapiro test fails reject the null hypothesis that the Indian stock returns follow a normal distribution at 1% level of significance and so the dynamics between the two markets are different. We also provide the Q-Q plot of the expected returns in Appendix H.

Instead of the variance, we propose to use the zscore rate of bankruptcy developed by Altman (1968) which forecast failure in the short-term. If a value less than 1.81 is returned, than there is a high probability of bankruptcy and if a value greater than 3.0 is returned, than there is a low probability of bankruptcy. We also use the expected risk premium per market risk β and design a performance measure that is useful when selecting stocks to be part of a portfolio.

2.3.2 Performance Measure

From the CAPM model, we can re-arrange and get:

$$\frac{E(r) - r_f}{\beta} = E(r_m) - r_f \tag{2.3}$$

Therefore, a stock performs well if

$$\frac{E(r) - r_f}{\beta} > E(r_m) - r_f \tag{2.4}$$

and vice versa.[11] However, among those that are doing well, we can distinguish between them based on their zscore of going bankrupt.[12] Now, Let $T = \frac{E(r) - r_f}{\beta}$ and $T_m = E(r_m) - r_f$, then we classify the performance of companies as Superior, Good and Poor. A company is classified as 'superior' if T is greater or equal to T_m and the zscore is higher than 3, 'good' if T is greater or equal to T_m and the zscore is less

[11]The left hand side expression is the Treynor's ratio.

[12]For detailed explanation on the zscore rate of bankruptcy, see Altman (1968).

or equal to 3 and 'poor' if T is less than T_m. The resulting performance measure is summarized in Figure I.1.

2.4 Empirical Models

2.4.1 Multinomial Logit Model

We specify the following form of the stock performance equation:

$$Y_i^* = X_i^{'}\beta + \varepsilon_i \tag{2.5}$$

Where i indexes the observations (companies), Y_i^* is the latent stock performance (unobserved), $X_i^{'}$ are the financial ratios affecting stock returns, $\varepsilon_i \sim G(0,1)$, Gumbel distribution (mutually independent). We follow Upadhyay et al. (2012) and estimate the model using the MNL model since we have three categories of performance.[13]

We do not know Y_i^* however we observe Y_i coded as 0,1,2 representing the three categories of performance (poor, good and superior). Thus, we have

$$Y_i = \begin{cases} 0 & \text{if the stock performs poor} \\ 1 & \text{if it performs average} \\ 2 & \text{if it has superior performance} \end{cases} \tag{2.6}$$

The normalized log-likelihood function of the samples is given by:

$$L_n(Y) = \frac{1}{n}\sum_{i}^{n}\sum_{j=0}^{2} D_{ij} lnPr\left(y_i = j|x_i\right) \tag{2.7}$$

[13] This is sometimes called multinomial logistic regression (MLR).

with

$$D_{ij} = \begin{cases} 1 & \text{if } y_i = j \\ \\ 0 & \text{if } y_i \neq j \end{cases} \tag{2.8}$$

and

$$P_{ij} = \frac{\exp X_i' \beta_j}{\sum_{m=0}^{2} \exp X_i' \beta_m} \tag{2.9}$$

Where $P_{ij} = Pr(y = j \mid x_i)$, $j = 0, 1, 2$ is the probability that an observation i belongs to alternative j and should be such that

$$\sum_{j=0}^{2} P_{ij} = 1, \forall i \tag{2.10}$$

The superior performance group is taken as the base category so the usual restriction is that $\beta_0 = 0$.

2.4.2 Artificial Neural Network

ANNs have enjoyed increasing popularity over standard econometric tools in predicting outcomes mainly because they do not suffer from specification bias and their ability to model highly complex relationships. A neural network model will be more complicated to explain than a regression model since the associated weights have no meaning. However, in most fields, management would prefer a stronger predictive model, even if it is more complicated. The predictive efficacy of ANNs have been shown in several fields, particularly in medicine and marketing.

Unlike ANNs, the coefficients in MNL model can be interpreted. However, when several ratios are put together and are correlated, it is difficult to assess the effects of the individual ratios just as ANNs. In addition, there are some underlying assumptions of MNL model which can have serious implications. Among them includes the assumption of 'independence from irrelevant alternatives' which reduces m alternatives to a series of pairwise comparisons that are unaffected by the characteristics of alternatives other than the pair under consideration. The second being the interaction effects between the explanatory variables which is difficult under MNL model.

Therefore, when MNL is being used solely for predictive purposes, it is important to ask how it compares with ANNs. Neural Networks have been increasingly applied to a wide range of finance problems, going from modelling financial markets (Refenes, 1995; Trippi and Turban, 1992) to loan risk analysis (Burgess, 1995). Bridle (1990) showed that the Softmax Output network with shared weights generalizes the MNL model and thus, we design a network which we believe can take into account the complex relationships in the stock market and compare the results to the MNL model.

We consider multi-layer feed forward network fully connected and consists of four layers. Each unit in a given layer is connected to every unit in the next layer and every connection has weight ($w_i j$) associated with it. The number of inputs in each of the hidden layers are determined internally using the testing dataset. The activation function is the sigmoid function given by:

$$f(X, W) = \frac{1}{1 + exp(-W^T X)} \tag{2.11}$$

The output function is given by

$$S_k = \frac{exp(W_k^T f(X, W))}{\sum_j exp(W_j^T f(X, W))} \tag{2.12}$$

where $j, k = 0, 1, 2$

The cross-entropy error function is given by

$$E(W) = -\sum_{i=1}^{N}\sum_{k=0}^{2} D_{ik} ln S_k(X_i, W) \tag{2.13}$$

where

$$D_{ik} = \begin{cases} 1 & \text{if } y_i = k \\ \\ 0 & \text{if } y_i \neq k \end{cases} \tag{2.14}$$

We will train the network with the scaled conjugate gradient algorithm with a batch training.[14] Details of the algorithm is standard and can be found in any textbook on neural networks.

2.5 Data

The dataset used for the empirical analysis comes from Compustat CD.[15] A sample period of 11 years starting from January, 2002 to December, 2012 is considered. For each year, the companies were selected based on two (2) pre-determined criteria. The company:

- must be public
- must actively be traded in the U.S.[16]

A total of 2133 companies are selected as a result of above criteria. Annual financial data on these companies have been extracted for each year and pooled

[14] We are in no means claiming this as being the best algorithm for the U.S. market. Our goal is to investigate any superiority of ANNs over MNL in the market and a such we will not be comparing the different algorithms in this work.

[15] Compustat is a database of financial, statistical and market information on companies throughout the world. The data come from the Compustat North America CD which focuses on US and Canada companies

[16] This excludes Canadian companies not trading on U.S. exchanges, but includes American Depository Receipts (ADRs).

together for two subperiods: 2011-2012 and 2005-2012. The use of the two sample periods have the ability to check the stability of the models over different time periods. [17] We have also collected quarterly data for the sample period 2011-2012 to investigate the influence of the data frequency.

The economic sector analyses have been conducted for three sectors: energy, industrials and informational technology using yearly data for sample period from 2002 to 2012. The final sample sizes used in this study are summarized in Table 2.1. We split the final sample into training and testing data. We build the models on the training data and test them using the testing data to help check over-fitting. In this work, we use 75% of the sample for training and 25% for testing. In some cases, we keep 5% of the testing data as a holdout for prediction. We have also collected information about Indian Stocks' returns from Yahoo Finance.

Table 2.1: Final Sample Size

	Sample Period	Sample Size
Yearly data	sample 1: (2011-2012)	1821
	sample 2: (2005-2012)	6667
Quarterly data	sample 3: (2011-2012)	5398
Industry yearly data	energy: (2002-2012)	952
	industrials: (2002-2012)	1895
	information technology: (2002-2012)	618

The Standard & Poor's 500 (S & P 500) stock index is used as a proxy to represent the U.S. market.[18] An annualized 6 monthly treasury bill rate is used for the yearly risk free rate while a 3 monthly t-bill is used as a proxy for the quarterly risk free rate. The size of each company is controlled for by their market capitalization and we standardized the values so that they are comparable with the ratios.

[17] In most cases, stocks selection are done at a particular point in time and therefore, we have to check how pooling different times together can influence the results.

[18] The stocks in the index are chosen for market size, liquidity, and industry group representation. It is the most commonly used benchmark for the overall US stock market.

According to economic theory, changes in financial ratios that affect changes in stock prices and dividend should effectively predict performance. There are more than 100 financial items that are reported in the literature (Chen and Shimerda, 1981).[19] In this paper, we have selected ratios which have empirical evidence of high predictive power on stock returns or measures the efficiency, liquidity, profitability, solvency and growth opportunities of firms.

We have identified 9 ratios to affect performance of the entire market and they are cash turnover (CT), current ratio (CURRENT), debt to market ratio (DMKT), financial leverage ratio (FLR), net profit margin (NPM), P/B, P/E, return on assets (ROA) and return on investments (ROI). The ratios hypothesized to be important in the energy sector are book value per share (BVPS), cash flow per share (CFS), CURRENT, DY and earnings per share (EPS). We propose BVPS, CFS, CURRENT, DY, EPS and ROA for the industrials sector and CT, CURRENT, DMKT, NPM and ROI for the informational technology sector. For basic sample statistics of these ratios, go to Appendix J.

2.6 Empirical Results

2.6.1 Yearly Data

The estimated results for the two different samples: 2011-2012 and 2005-2012 are presented in Table K.1 and Table K.2. In general, having more data did not change the signs of the logistic coefficients and where there was a change, the coefficient was not statistically significant. The size of the companies measured by their standardized market capitalization tends out not to be significant in distinguishing the alternative performance groups from the reference group (superior) in both samples. A firm's

[19] Chen and Shimerda (1981) reported that more than 100 financial indicators have been analyzed in 26 studies, of which 65 are accounting ratios. 41 of these indicators are considered useful and/or are used in the final analysis by one or more authors.

efficiency in its use of cash for generation of sales revenue as measured by CT is very significant and decreases the log odds by 0.001 in both samples.

CURRENT is significant in distinguishing between the superior and good performance groups but not the superior group from the poor performing group. CURRENT measures a firm's ability to settle its short term debt obligations. A firm's inability to pay its debts will mean going bankrupt, captured by the zcore which was used to separate the superior and the good and thus consistent with how performance was created. The DMKT is highly significant and distinctive among the groups in both samples.

The overall relationship among the ratios and the stock performance groups was assessed by a likelihood ratio test and there is a significance evidence that the coefficients are all jointly different from zero. We now assess the effect of each hypothesized ratio in the model using likelihood ratio test. We run a reduced model by omitting an effect from the final model and test the null hypothesis that all coefficients of that effect are zero and report the p-values in Table L.1. Surprisingly, P/E shown by previous research as having strong effect on stock return is not significant.

As mentioned earlier, we want to find the right financial ratios and model for prediction and thus we will study the classification accuracy table from the MNL. The table is created by comparing the predicted categories based on the model against the observed response groups in order to assess the prediction power of the model. Even if the financial ratios have no relationship with stock performance, one would expect some of the groups to be predicted correct by chance. This is referred to as 'by chance accuracy' and it is computed by summing the squared proportion of the observations in each category. The benchmark that is used to characterize a MNL as useful is a 25% improvement over the chance accuracy rate.[20] Figure M.1 compares the 'by accuracy rate' and the 'benchmark rate' for the different samples we considered for

[20]We will refer to this as the 'benchmark rate'.

this study.

Now that we know what prediction rate would constitute a good model, we will present the results from the models. We build the model using the training data, predict the training data and then validates the model with the testing data. The results are shown in Table 2.2. The overall prediction rate for the sample: 2011-2012 is the best one would expect for an efficient market like NYSE and far exceeds the rate obtained for the Indian market by Upadhyay et al. (2012). The 63.4% accuracy rate is greater than the benchmark rate of 58.2% and less than the accuracy rate 65.1% obtained from the testing data, thus suggesting a good model. However, the results obtained for the sample: 2005-2012 suggest over-fitting the model since the prediction rate from the testing data less that in the training even though the obtained accuracy rate is higher than the benchmark rate.

Table 2.2: Classification Table (Yearly Data)

		Percentage Correct (%)	
Data	Performance	2011-2012	2005-2012
Training	Poor	93.0	78.1
	Good	3.9	4.7
	Superior	20.3	44.7
	Overall	63.4	50.9
Testing	Poor	93.7	73.9
	Good	0	5.9
	Superior	18.3	43.6
	Overall	65.1	48.8

Due to the inconsistency of results from the MNL, we run goodness-of-fit tests under the null hypothesis that the model adequately fits the data. We computes Pearson and Deviance goodness-of-fit statistics so that we can compare our results here to Upadhyay et al. (2012) found for the Indian market. Under the null, the Pearson and Deviance statistics have a chi-square distribution. The results displayed in Table N.1 contradict each other: while the Deviance statistic fails to reject the null hypothesis, Pearson statistic finds evidence that the model does not adequately fit the data in both samples with p-value = 0. It is therefore difficult for one to completely trust the MNL as a predictive tool for the U.S. market.

2.6.2 Quarterly Data

One hypothesis in this study is whether the frequency of data matters and we have only done the analyses up to a quarterly data. The estimated results as shown in Table K.3, however we see that they are qualitatively about the same as those obtained from the yearly data. The individual ratios' effects are also presented in Table L.2. Except P/E, all the hypothesized ratios significantly contribute to the model and therefore, the size of the company is now important at a lower data frequency.

Table 2.3 shows the prediction accuracy rates from the training and testing quarterly data.

Table 2.3: Classification Table (Quarterly Data)

	Percentage Correct (%)	
Data	Performance	2011-2012
Training	Poor	43.5
	Good	41.6
	Superior	76.4
	Overall	55.6
Testing	Poor	45.0
	Good	42.9
	Superior	75.5
	Overall	56.3

2.6.3 ANNs

The number of units in each hidden layer is determined optimally using a testing data set. A network with eight hidden units on the first hidden layer and six hidden units on the second hidden layer was selected as a result all samples. An example of such network is shown in Figure E.1. In general, we conclude that the ANNs is able to model the complex relationships in the U.S. data better as seen in the classification table in Table 2.4. The network is consistent with a higher overall prediction accuracy rate in the testing phase in all samples suggesting no over-fitting. The network is able

Table 2.4: Classification Table (ANNs)

		Percentage Correct (%)		
Data	Performance	2011-2012	2011-2012(Q)	2005-2012
Training	Poor	95.7	45.4	63.8
	Good	2.5	47.8	26.7
	Superior	19.0	73.1	55.2
	Overall	64.3	56.5	52.1
Testing	Poor	92.2	46.9	63.4
	Good	1.7	47.2	31.2
	Superior	24.7	73.1	58.1
	Overall	66.5	57.3	54.5

to capture the inter-relationships existing among the ratios which otherwise cannot be achieved using MNL. Achieving more than 50% overall classification rate in an efficient market like the U.S. market is a success but the synaptic weights cannot be interpreted. As we have seen in this study, several set of ratios need to be put together to obtain a higher predictive rate. In this case, the marginal effects of the ratios would be difficult to interpret since the ratios may be correlated. Therefore, when prediction is all we care about, this work proposes ANNs over MNL.

2.6.4 Industry Sector Analyses

For the entire U.S. market, we have shown that MNL is always rejected by Pearson goodness-of-fit test as a good model for probabilistic prediction and could potentially lead to over-fitting. This study further shows that these inconsistencies can be eliminated and finds that, MNL passes all the goodness-of-fit tests at the industry level.

Evidence from three industries (Energy, Industrials and Information technology) are discussed in Sections 2.6.4.1– 2.6.4.3. We also find MNL to be more predictive in the information technology industry compared to the other industries considered.

2.6.4.1 Energy

We have identified five important financial ratios in the energy industry: BVPS, CFS, CURRENT, DY and EPS as key in predicting performance in a probabilistic framework. The coefficients with their p-values are reported in Table K.4. The various goodness-of-fit tests, we run in this work shows MNL fits the data (refer to Table L.3 and Table N.3).

We also check the predictive power of this model in the energy sector and report the classification table in Table 2.5. The overall accuracy rate at both the training and testing phases is higher than the benchmark rate (see Figure M.2) also suggesting a good model but the rate is small as shown in Table 2.5 and we suspect this could be the fact that the energy sector responds to information more quickly and thus more efficient relative to the whole market or the ratios are not enough and other factors are needed to improve the prediction rate.

Table 2.5: Classification Table (Energy)

	Percentage Correct (%)	
Data	Performance	2002-2012
Training	Poor	80.7
	Good	8.8
	Superior	36.7
	Overall	48.5
Testing	Poor	82.6
	Good	6.9
	Superior	39.1
	Overall	48.9

2.6.4.2 Industrials

In addition to the set of ratios identified for the energy sector, ROA is needed to make the MNL model works for the industrial sector. Pearson and Deviance goodness-of-tests failed to find evidence against the model at 5% level of significance (see Table N.3). The significance of the individuals' proposed ratios using likelihood ratio test are provided in Table L.3. While BVPS is not significant in distinguishing the good and poor performance from the superior group as seen displayed in Table K.5, ROA is highly significant. CURRENT on the other hand is significant in the good group but not the poor which is consistent with how we created the performance group.

Now, we look at the prediction rate in this sector. Table 2.6 shows accuracy rates in a training and testing sample. The rates achieved are about the same as the rates obtained for the energy sector and again, revealing another similarity between these two sectors. The rates are higher than the benchmark rate and the MNL model can be used for prediction in the probabilistic framework.

Table 2.6: Classification Table (Industrials)

	Percentage Correct (%)	
Data	Performance	2002-2012
Training	Poor	74.4
	Good	11.2
	Superior	40.9
	Overall	48.2
Testing	Poor	75.6
	Good	6.8
	Superior	43.9
	Overall	49.3

2.6.4.3 Information Technology

MNL is most predictive in the information technology sector. This could be the fact that the information technology sector is less efficient compared to the energy and industrials sectors or we have identified ratios that are key determinant in this industry. Likelihood ratio test (Table L.3) reveals CT, CURRENT and DMKT as significant factors. The p-values for Pearson and Deviance goodness-of-fit tests are 0.994 and 0.998 respestively and we thus fail to reject the null hypothesis that the MNL model fits the data.

The overall accuracy rates are 56.2% and 56.5% in training and testing data respectively as seen in Table 2.7. The probability of a correct classification is 0.33 and thus achieving there accuracy rates which are also greater than the benchmark rate of 53.11% (see Figure M.2) suggests that the MNL model works for this industry with proposed ratios.

Table 2.7: Classification Table (Information Technology)

	Percentage Correct (%)	
Data	Performance	2002-2012
Training	Poor	87.3
	Good	0.0
	Superior	23.7
	Overall	56.2
Testing	Poor	89.9
	Good	14.3
	Superior	14.7
	Overall	56.5

2.7 Conclusion and Future Research

This paper has introduced a probabilistic prediction of stock performance in the U.S. market, an efficient market where one would expect a zero alpha and showed a potential prediction possibilities. MNL is always rejected by Pearson goodness-of-fit test irrespective of the sample periods and the data frequency considered for this study. Previous studies analyzing activities in the U.S. stock markets have relied heavily on DY, EPS and P/E. This study demonstrates that other ratios such as CT, CURRENT, DMKT and ROA play a key role on a company's stock performance.

We have seen that MNL fits the U.S. data at the industry level but the peudo R-squares are small suggesting that there may be other important factors affecting stock returns and thus performance. Further research into other economic and industry specific factors to help improve the prediction rate is strongly encouraged. Where industry specific factors exist, nested logit model can be implemented to see how it fits the data as well as its predictive power.

This paper shows that a neural network is preferred when prediction is the ultimate goal. By no means are we claiming that this is the best prediction rate one can derive from the network. ANNs are may be highy unstable: we have seen that the order of the ratios can seriously affect the prediction rate and so one may ask what is the maximum prediction rate from the network? Which algorithm among the numerous algorithms available is optimum? Future studies will be interesting to answer these questions.

APPENDICES

APPENDIX A

Derivation of the Bühlmann Straub Credibility Model

Defining the number of claims per policy as $\frac{N_{ij}}{\lambda_{ij}}$ and by our assumption of the distribution of the number of claims, we have that

$$\begin{aligned} E(F_{ij}|\Theta_i) &= E\left(\frac{N_{ij}}{\lambda_{ij}} \mid \Theta_i\right) = \mu(\Theta_i) \\ Var(F_{ij}|\Theta_i) &= Var\left(\frac{N_{ij}}{\lambda_{ij}} \mid \Theta_i\right) = \frac{\sigma^2(\Theta_i)}{\lambda_{ij}} \end{aligned} \tag{A.1}$$

- Now, define

$$\widehat{\mu(\Theta_i)} = a_i + \sum_{j=1}^{J} b_{ij} F_{ij} = a_i + b_i \sum_{j=1}^{J} F_{ij} \tag{A.2}$$

where the last equality results from the probability distribution of $F_{i1}, \cdots, F_{iJ}$ being invariant under the permutations of F_{ij} and the uniqueness of the credibility estimator (Bühlmann and Gisler, 2005).

Lemma A.1. *(Bühlmann, 1967) If $E(a_i + b_i F_i - \mu(\theta))^2 \leq E(a'_i + b'_i F_i - \mu(\theta))^2$ for all arbitrary a' and b',*

then $a_i + b_i F_i$ is also the best linear approximation to $E(\mu(\theta)|F_{i1}, F_{i2}, ..., F_{iJ})$. [1]

[1]Refer to (Bühlmann, 1967) for the proof of the lemma.

- Now, by letting $F_i = \sum_{j=1}^{J} \frac{\lambda_{ij}}{\lambda_i} F_{ij}$ and $\lambda_i = \sum_{j=1}^{J} \lambda_{ij}$

$$\widehat{\mu(\Theta_i)} = \hat{a}_i + \hat{b}_i F_i \tag{A.3}$$

Where

$$(\hat{a}_i, \hat{b}_i) = \arg\min_{a_i, b_i} E\left(\mu(\Theta_i) - a_i - b_i F_i\right)^2 \tag{A.4}$$

- The first order conditions yield

$$\begin{aligned} &E(\mu(\Theta_i) - a_i - b_i F_i) = 0 \\ &Cov(\mu(\Theta_i), F_i) = b_i Var(F_i) \end{aligned} \tag{A.5}$$

- After some simplifications and combining equations (A.3) and (A.5) , we get

$$\widehat{\mu(\Theta_i)} = E\mu(\Theta_i) + [Var(\mu(\Theta_i))]^2 \left(\frac{E\sigma^2(\Theta_i)}{\lambda_i Var(\mu(\Theta_i))} + 1\right)^{-1} (F_i - EF_i) \tag{A.6}$$

APPENDIX B

Moments of Weibull Count Model

The expected value and variance of the Weibull count model, given a heterogeneous rate parameter Θ_i are

$$\mu(\Theta_i) = \sum_{n=1}^{\infty}\sum_{m=n}^{\infty} \frac{n(-1)^{m+n}\alpha_m^n}{\Gamma(cm+1)}\Theta_i^m$$

$$\sigma^2(\Theta_i) = \sum_{n=1}^{\infty}\sum_{m=n}^{\infty} \frac{n^2(-1)^{m+n}\alpha_m^n}{\Gamma(cm+1)}\Theta_i^m - \left(\sum_{n=1}^{\infty}\sum_{m=n}^{\infty} \frac{n(-1)^{m+n}\alpha_m^n}{\Gamma(cm+1)}\Theta_i^m\right)^2 \tag{B.1}$$

Where $\alpha_m^0 = \frac{\Gamma(cm+1)}{\Gamma(m+1)}$, for $m = 0, 1, 2, ...$ and $\alpha_m^{n+1} = \sum_{l=n}^{m-1} \frac{\Gamma(cm-cl+1)}{\Gamma(m-l+1)}$, for $n = 0, 1, 2, ...$ for $m = n+1, n+2, n+3, ...$

APPENDIX C

Structural Parameters: Weibull Count Model

The derivation of the heterogeneous Weibull count model when t = 1 is given by

$$
\begin{aligned}
P(N=n) &= \int_0^\infty \left[\sum_{m=n}^{\infty} \frac{(-1)^{m+n}\alpha_m^n \Theta_i^m}{\Gamma(cm+1)}\right] g(\Theta_i|r,\alpha)d\Theta_i \\
&= \int_0^\infty \left[\sum_{m=n}^{\infty} \frac{(-1)^{m+n}\alpha_m^n \Theta_i^m}{\Gamma(cm+1)}\right] \times \frac{\alpha^r(\Theta_i)^{r-1}e^{-\alpha\Theta_i}}{\Gamma(r)}\mathrm{d}\Theta_i \\
&= \sum_{m=n}^{\infty} \frac{(-1)^{m+n}\alpha_m^n}{\Gamma(cm+1)} \int_0^\infty \Theta_i^m \frac{\alpha^r(\Theta_i)^{r-1}e^{-\alpha\Theta_i}}{\Gamma(r)}\mathrm{d}\Theta_i \\
&= \sum_{m=n}^{\infty} \frac{(-1)^{m+n}\alpha_m^n}{\Gamma(cm+1)} \int_0^\infty \frac{(\alpha\Theta_i)^{m+r-1}e^{-\alpha\Theta_i}}{\Gamma(r)\alpha^m}\mathrm{d}\alpha\Theta_i \\
&= \sum_{m=n}^{\infty} \frac{(-1)^{m+n}\alpha_m^n}{\Gamma(cm+1)} \frac{1}{\Gamma(r)\alpha^m} \int_0^\infty (\alpha\Theta_i)^{m+r-1}e^{-\alpha\Theta_i}\mathrm{d}\alpha\Theta_i \\
&= \sum_{m=n}^{\infty} \frac{(-1)^{m+n}\alpha_m^n}{\Gamma(cm+1)} \frac{\Gamma(r+m)}{\Gamma(r)\alpha^m}
\end{aligned}
\tag{C.1}
$$

For $n = 0, 1, 2, ...$

The collective premium μ_0 is the expected value of the individual risk premiums given by

$$
\mu_0 = E[\mu(\Theta_i)] = \sum_{n=1}^{\infty}\sum_{m=n}^{\infty} \frac{n(-1)^{m+n}\alpha_m^n E(\Theta_i^m)}{\Gamma(cm+1)} \tag{C.2}
$$

Where the last equality holds because of the linearity property of expectation. Therefore, we have that

$$\begin{aligned}
\hat{\mu}_0 &= \sum_{n=1}^{\infty}\sum_{m=n}^{\infty} \frac{n(-1)^{m+n}\alpha_m^n \widehat{E(\Theta_i^m)}}{\Gamma(\hat{c}m+1)} \\
&= \sum_{n=1}^{\infty}\sum_{m=n}^{\infty} \frac{n(-1)^{m+n}\alpha_m^n}{\Gamma(\hat{c}m+1)} \frac{\Gamma(\hat{r}+m)}{\Gamma(\hat{r})\hat{\alpha}^m} \\
&= \sum_{n=1}^{\infty}\sum_{m=n}^{\infty} \frac{n(-1)^{m+n}\alpha_m^n}{\Gamma(\hat{c}m+1)} \frac{(\hat{r}+m-1)...\hat{r}}{\hat{\alpha}^m}
\end{aligned} \tag{C.3}$$

Since $\Theta_i \sim gamma(r, \alpha)$ and $E(\Theta_i^m) = \frac{\Gamma(r+m)}{\Gamma(r)\alpha^m}$, which is the mth moment of the gamma distribution.

The expected variance within individual risks σ^2 is equal to

$$\sigma^2 = E(\sigma^2(\Theta_i)) = \sum_{n=1}^{\infty}\sum_{m=n}^{\infty} \frac{n^2(-1)^{m+n}\alpha_m^n}{\Gamma(cm+1)} (E(\Theta_i))^m - \widehat{E}\left(\sum_{n=1}^{\infty}\sum_{m=n}^{\infty} \frac{n(-1)^{m+n}\alpha_m^n}{\Gamma(\hat{c}m+1)}\Theta_i^m\right)^2 \tag{C.4}$$

Using the best estimator for $E(\Theta_i)$ and the maximum likelihood estimates, we obtain the expression in equation (1.9).

Lastly, we will find an estimate for the variance between individual risk premiums using the mean expression of the Weibull count model. This is given by

$$\tau^2 = Var(\mu(\Theta_i)) = Var\left(\sum_{n=1}^{\infty}\sum_{m=n}^{\infty} \frac{n(-1)^{m+n}\alpha_m^n}{\Gamma(cm+1)}\Theta_i^m\right) \tag{C.5}$$

Now, using the maximum likelihood estimates, $\hat{\tau}^2$ is equal to equation (1.10).

APPENDIX D

Monte Carlo Estimates

Let

$$f(\Theta_i) = \sum_{n=1}^{\infty} \sum_{m=n}^{\infty} \frac{n(-1)^{m+n} \alpha_m^n}{\Gamma(\hat{c}m + 1)} \Theta_i^m \tag{D.1}$$

Now, take n-sample of Θ_i, $(\theta_{i1}, \cdots, \theta_{in})$, then an unbiased estimate of $E(f(\Theta_i))$ is

$$\tilde{f}_n(\Theta_i) = \frac{1}{n} \sum_{h=1}^{n} f(\theta_{ih}) \tag{D.2}$$

Similarly, an unbiased estimate of $Var(f(\Theta_i))$ is given by:

$$\widehat{Var}(f(\Theta_i)) = \frac{1}{n-1} \sum_{h=1}^{n} \Big(f(\theta_{ih}) - \tilde{f}_n(\theta_i) \Big)^2 \tag{D.3}$$

The proof of unbiasedness of the expected value is

$$E(\tilde{f}_n(\theta_i)) = E\left(\frac{1}{n} \sum_{h=1}^{n} f(\theta_{ih}) \right) = \frac{1}{n} \sum_{h=1}^{n} E(f(\theta_{ih})) = E(f(\Theta_i)) \tag{D.4}$$

APPENDIX E

Simulation Results

Table E.1: $r = 5, \alpha = 4, c = 2$: (Mean, Median)

True Parameter	(I=3500, n=137)	(I=5000, n=131)	(I=8000, n=137)	(I=10000, n=145)
$r = 5$	4.861, 4.868	4.843, 4.840	4.857, 4.848	4.871, 4.863
$\alpha = 4$	3.829, 3.819	3.840, 3.832	3.841, 3.833	3.845, 3.830
$c = 2$	2.016, 2.017	2.016, 2.015	2.013, 2.012	2.006, 2.008

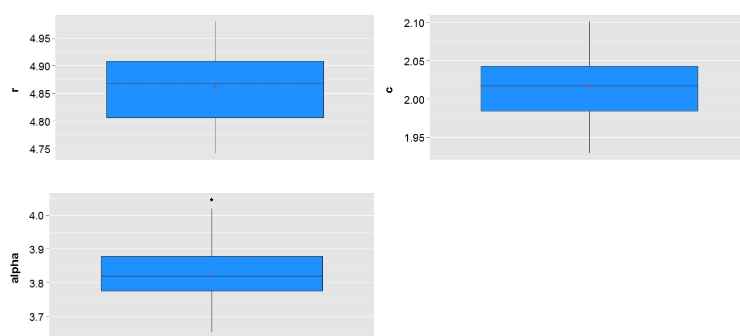

Figure E.1: I=3500, n=137

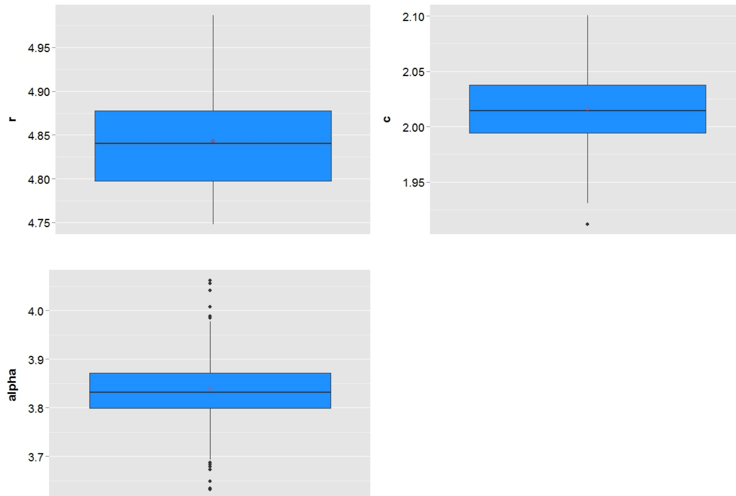

Figure E.2: I=5000, n=131

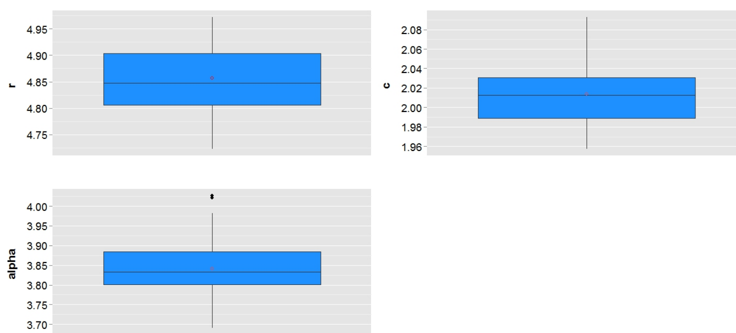

Figure E.3: I=8000, n=137

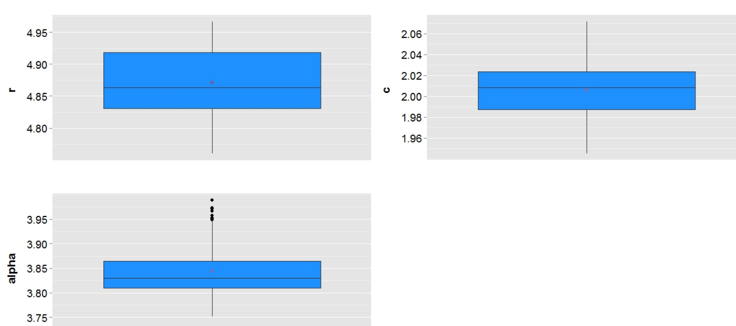

Figure E.4: I=10000, n=145

Table E.2: $r = 7, \alpha = 3, c = 1$: (Mean, Median)

True Parameter	(I=3500, n=29)	(I=5000, n=35)	(I=8000, n=39)	(I=10000, n=48)
$r = 7$	6.705, 6.700	6.688, 6.690	6.686, 6.691	6.701, 6.704
$\alpha = 3$	2.837, 2.838	2.843, 2.844	2.848, 2.845	2.837, 2.839
$c = 1$	1.018, 1.019	1.012, 1.011	1.014, 1.013	1.012, 1.011

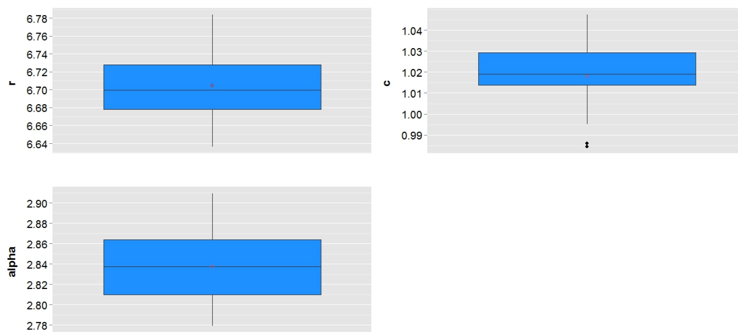

Figure E.5: I=3500, n=29

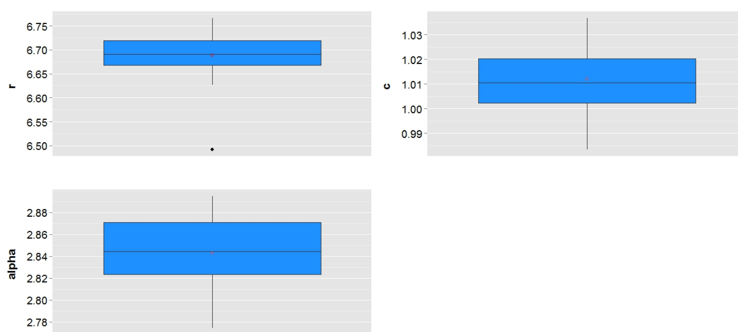

Figure E.6: I=5000, n=35

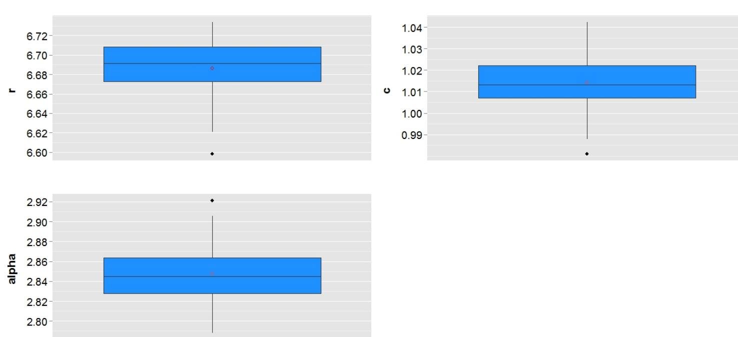

Figure E.7: I=8000, n=39

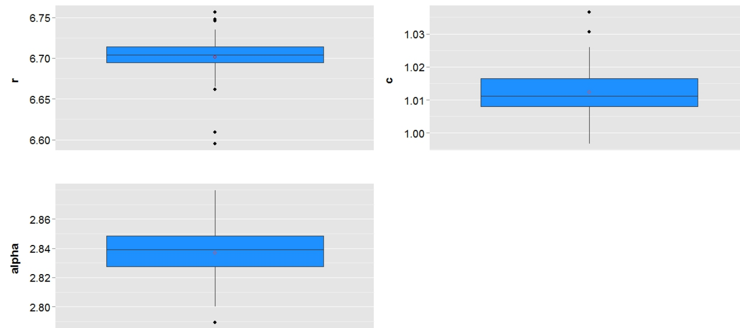

Figure E.8: I=10000, n=48

Table E.3: $r = 7, \alpha = 3, c = 0.95$: (Mean, Median)

True Parameter	(I=3500, n=88)	(I=5000, n=52)	(I=8000, n=47)	(I=10000, n=54)
$r = 7$	6.952, 6.998	6.879, 6.904	6.968, 6.913	6.904, 6.788
$\alpha = 3$	2.976, 2.995	2.942, 2.948	2.962, 2.946	2.975, 2.936
$c = 0.95$	0.950, 0.947	0.952, 0.949	0.958, 0.956	0.949, 0.948

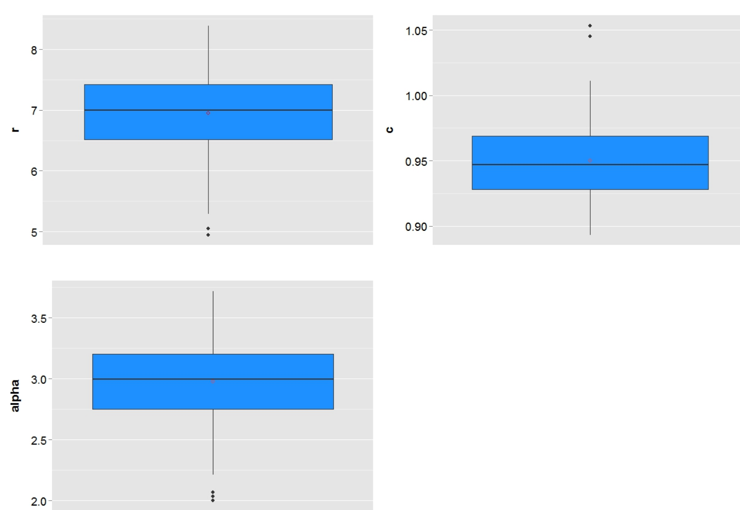

Figure E.9: I=3500, n=88

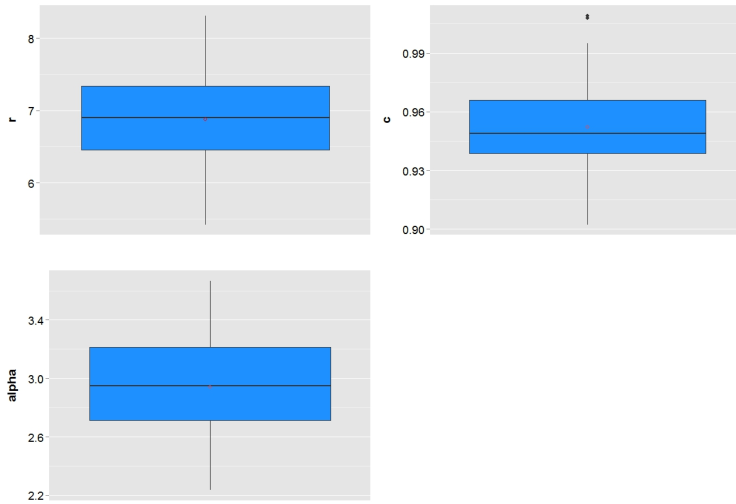

Figure E.10: I=5000, n=52

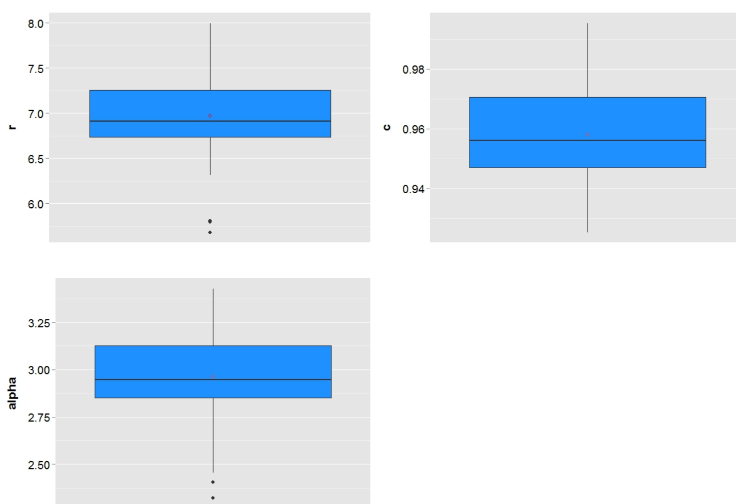

Figure E.11: I=8000, n=47

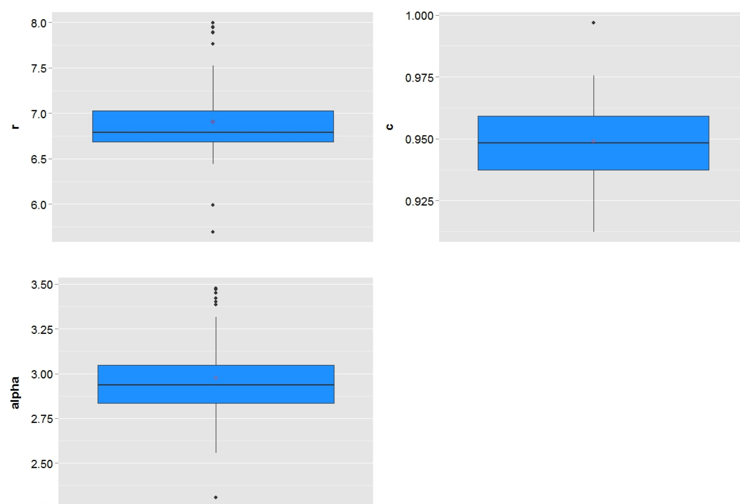

Figure E.12: I=10000, n=54

APPENDIX F

Moments of Mittag Count Model

The expected value and variance of the Mittag Leffler count model, given a heterogeneous rate parameter Θ_i are

$$\begin{aligned}\mu(\Theta_i) &= \frac{\Theta_i^c}{\Gamma(1+c)} \\ \sigma^2(\Theta_i) &= \frac{\Theta_i^c}{\Gamma(1+c)} + \frac{2\Theta_i^{2c}}{\Gamma(1+2c)} - \left(\frac{\Theta_i^c}{\Gamma(1+c)}\right)^2\end{aligned} \tag{F.1}$$

APPENDIX G

Structural Parameters: Mittag Leffler Count Model

The derivation of the heterogeneous Mittag Leffler count model when t = 1 is given by

$$
\begin{aligned}
P(N=n) &= \int_0^\infty \left[\sum_{m=n}^\infty \frac{(-1)^{m-n}\binom{m}{n}\Theta_i^{cm}}{\Gamma(cm+1)}\right] g(\Theta_i|r,\alpha)d\Theta_i \\
&= \int_0^\infty \left[\sum_{m=n}^\infty \frac{(-1)^{m-n}\binom{m}{n}\Theta_i^{cm}}{\Gamma(cm+1)}\right] \times \frac{\alpha^r(\Theta_i)^{r-1}e^{-\alpha\Theta_i}}{\Gamma(r)}\mathrm{d}\Theta_i \\
&= \sum_{m=n}^\infty \frac{(-1)^{m-n}\binom{m}{n}}{\Gamma(cm+1)} \int_0^\infty \Theta_i^{cm}\frac{\alpha^r(\Theta_i)^{r-1}e^{-\alpha\Theta_i}}{\Gamma(r)}\mathrm{d}\Theta_i \\
&= \sum_{m=n}^\infty \frac{(-1)^{m-n}\binom{m}{n}}{\Gamma(cm+1)} \int_0^\infty \frac{\alpha^r(\Theta_i)^{cm+r-1}e^{-\alpha\Theta_i}}{\Gamma(r)}\mathrm{d}\Theta_i \\
&= \sum_{m=n}^\infty \frac{(-1)^{m-n}\binom{m}{n}}{\Gamma(cm+1)} \int_0^\infty \frac{(\alpha\Theta_i)^{cm+r-1}e^{-\alpha\Theta_i}}{\Gamma(r)\alpha^{cm}}\mathrm{d}\alpha\Theta_i \\
&= \sum_{m=n}^\infty \frac{(-1)^{m-n}\binom{m}{n}}{\Gamma(cm+1)} \frac{1}{\Gamma(r)\alpha^{cm}} \int_0^\infty (\alpha\Theta_i)^{cm+r-1}e^{-\alpha\Theta_i}\mathrm{d}\alpha\Theta_i \\
&= \sum_{m=n}^\infty \frac{(-1)^{m-n}\binom{m}{n}}{\Gamma(r)\alpha^{cm}} \frac{\Gamma(cm+r)}{\Gamma(cm+1)}
\end{aligned}
\tag{G.1}
$$

For $n = 0, 1, 2, ...$

The collective premium μ_0 is the expected value of the individual risk premiums given by

$$\begin{aligned}
\mu_0 = E(\mu(\Theta_i)) = E\left(\frac{\Theta_i^c}{\Gamma(1+c)}\right) &= \frac{1}{\Gamma(1+c)}E(\Theta_i^c) \\
&= \frac{1}{\Gamma(1+c)}\frac{\Gamma(r+c)}{\Gamma(r)\alpha^c} \\
&= \frac{\Gamma(r+c)}{\Gamma(1+c)\Gamma(r)\alpha^c}
\end{aligned} \tag{G.2}$$

Therefore,

$$\widehat{\mu_0} = \frac{\Gamma(\hat{r}+\hat{c})}{\Gamma(1+\hat{c})\Gamma(\hat{r})\hat{\alpha}^{\hat{c}}} \tag{G.3}$$

Since $\Theta_i \sim gamma(r, \alpha)$ and $E(\Theta_i^c) = \frac{\Gamma(r+c)}{\Gamma(r)\alpha^c}$, which is the cth moment of the gamma distribution.

The expected variance within individual risks σ^2 is equal to

$$\begin{aligned}
\sigma^2 = E(\sigma(\Theta_i)) &= E\left(\frac{\Theta_i^c}{\Gamma(1+c)} + \frac{2\Theta_i^{2c}}{\Gamma(1+2c)} - \frac{\Theta_i^{2c}}{\Gamma(1+c)^2}\right) \\
&= \frac{E(\Theta_i^c)}{\Gamma(1+c)} + \frac{2E(\Theta_i^{2c})}{\Gamma(1+2c)} - \frac{E(\Theta_i^{2c})}{\Gamma(1+c)^2} \\
&= \frac{\Gamma(r+c)}{\Gamma(1+c)\Gamma(r)\alpha^c} + \frac{2\Gamma(r+2c)}{\Gamma(1+2c)\Gamma(r)\alpha^{2c}} - \frac{\Gamma(r+2c)}{\Gamma(1+c)^2\Gamma(r)\alpha^{2c}} \\
&= \frac{1}{\Gamma(r)\alpha^c}\left(\frac{\Gamma(r+c)}{\Gamma(1+c)} + \frac{2\Gamma(r+2c)}{\Gamma(1+2c)\alpha^c} - \frac{\Gamma(r+2c)}{\Gamma(1+c)^2\alpha^c}\right)
\end{aligned} \tag{G.4}$$

Using the MLEs, we obtain the expression in equation (1.17).

Lastly, we will find an estimate for the variance between individual risk premiums using the mean expression of the Mittag Leffler count model. This is given by

$$\begin{aligned}
\tau^2 = Var(\mu(\Theta_i)) &= Var\left(\frac{\Theta_i^c}{\Gamma(1+c)}\right) \\
&= E\left(\frac{\Theta_i^c}{\Gamma(1+c)}\right)^2 - \left(E\left(\frac{\Theta_i^c}{\Gamma(1+c)}\right)\right)^2 \\
&= E\left(\frac{\Theta_i^{2c}}{\Gamma(1+c)^2}\right) - \left(\frac{\Gamma(r+c)}{\Gamma(1+c)\Gamma(r)\alpha^c}\right)^2 \\
&= \frac{\Gamma(r+2c)}{\Gamma(1+c)^2\Gamma(r)\alpha^{2c}} - \frac{\Gamma(r+c)^2}{\Gamma(1+c)^2\Gamma(r)^2\alpha^{2c}} \\
&= \frac{1}{\Gamma(1+c)^2\Gamma(r)\alpha^{2c}}\left(\Gamma(r+2c) - \frac{\Gamma(r+c)^2}{\Gamma(r)}\right)
\end{aligned} \tag{G.5}$$

Again, using the MLEs, $\hat{\tau}^2$ is equal to equation (1.18).

APPENDIX H

Distribution of Stock Returns

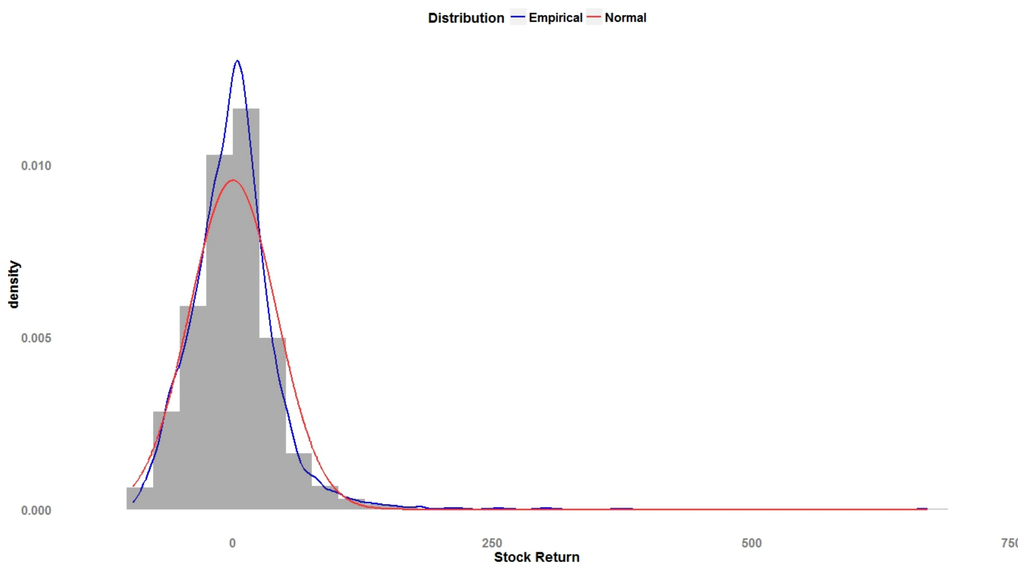

Figure H.1: Histogram of Stock Returns (INDIA)

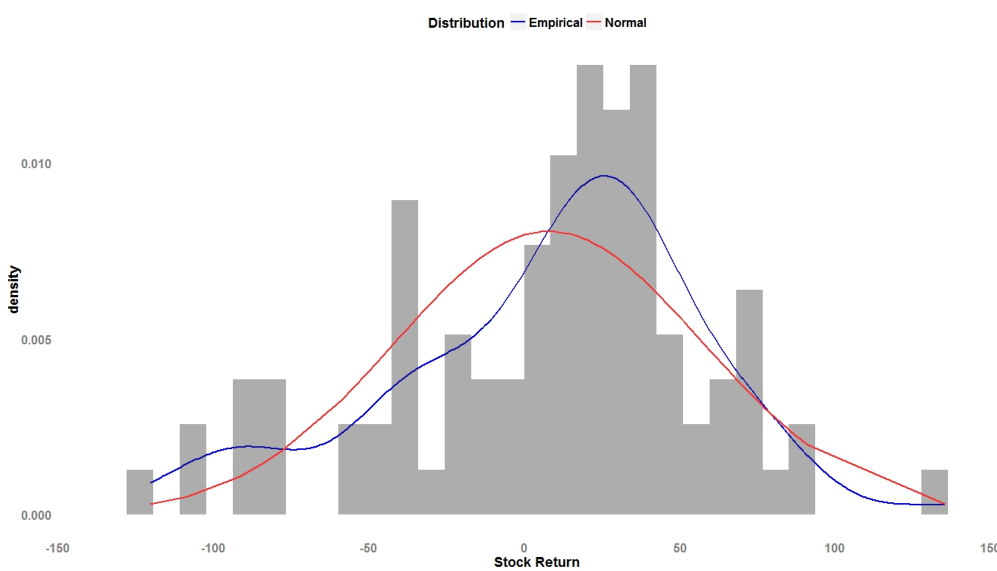

Figure H.2: Histogram of Stock Returns (NYSE)

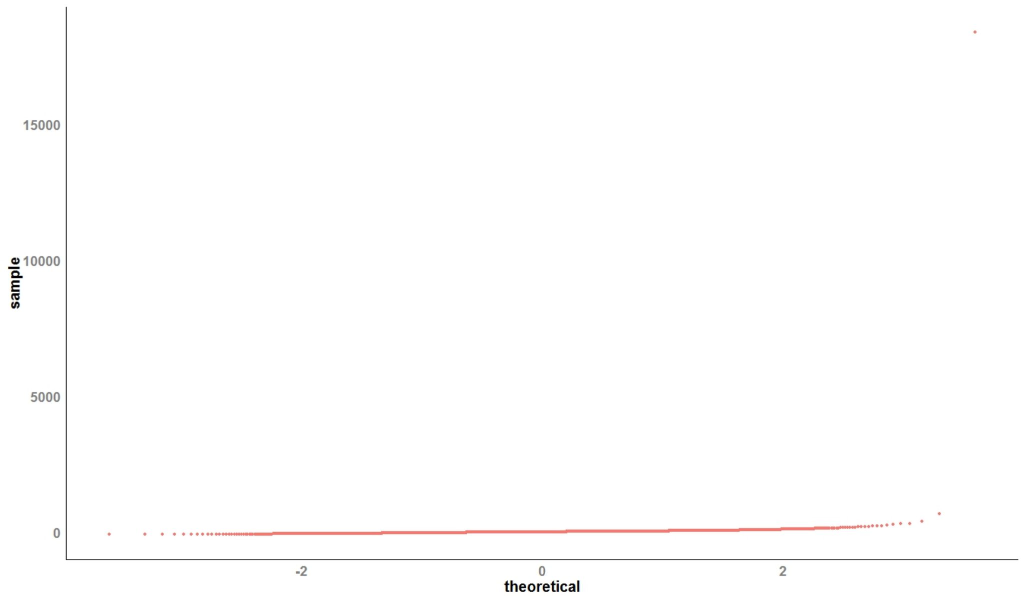

Figure H.3: Q-Q Plot (NYSE)

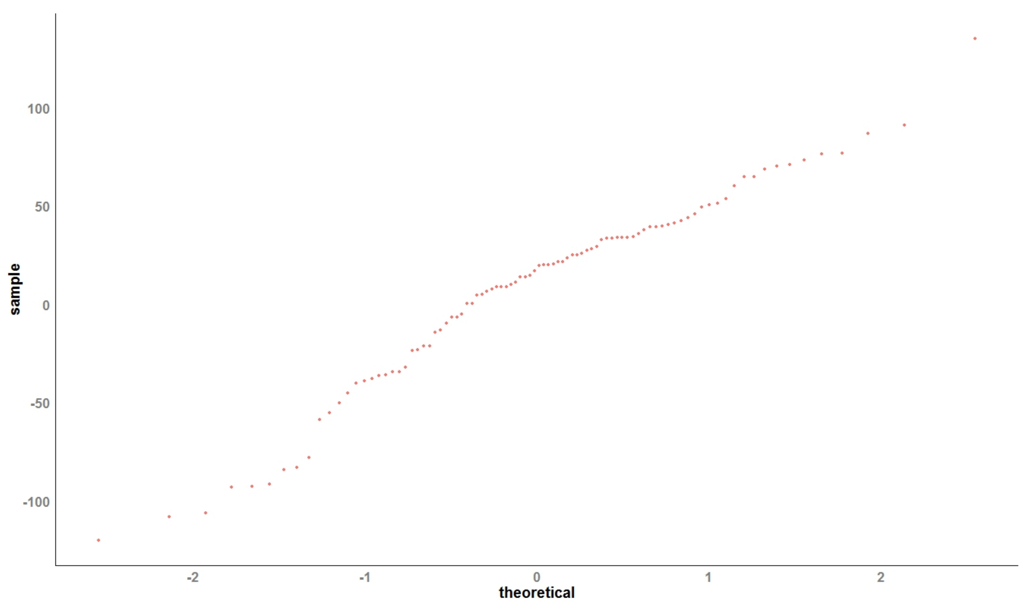

Figure H.4: Q-Q Plot (INDIA)

APPENDIX I

Performance Tree

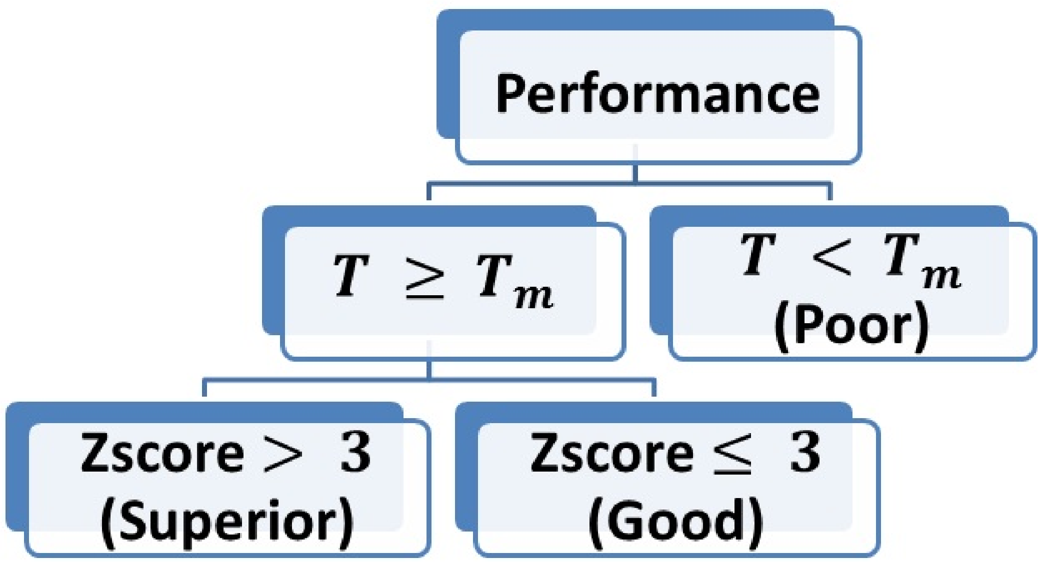

Figure I.1: Performance Tree

APPENDIX J

Sample Descriptive Statistics

Table J.1: Summary Statistics (Yearly Data: 2011-2012)

Ratios	Mean	St. Dev.	Min	Median	Max
CT	41.060	170.285	0.123	11.758	3,166.339
CURRENT	2.118	1.400	0.196	1.790	19.227
DMKT	0.547	0.983	0.000	0.298	16.684
FLR	4.633	45.767	1.077	2.362	1,918.410
NPM	4.335	32.664	−1,117.231	5.890	68.188
P/B	3.866	26.023	0.143	1.904	765.740
P/E	15.730	76.677	−1,158.333	15.283	1,316.000
ROA	4.643	8.532	−143.768	4.866	72.242
ROI	7.221	14.378	−166.021	7.299	142.464

Table J.2: Summary Statistics (Yearly Data: 2005-2012)

Ratios	Mean	St. Dev.	Min	Median	Max
CT	41.009	130.502	0.012	13.831	3,203.112
CURRENT	2.049	1.509	0.102	1.735	27.532
DMKT	0.516	1.171	0.000	0.259	44.288
FLR	3.636	25.175	1.022	2.328	1,918.410
NPM	−1.290	404.912	−31,688.900	5.820	4,251.798
P/B	3.395	18.475	0.060	2.009	922.476
P/E	15.078	115.621	−3,657.001	15.729	4,241.001
ROA	4.472	12.747	−678.887	5.046	84.047
ROI	6.932	20.581	−943.228	7.639	142.464

Table J.3: Summary Statistics (Quarterly Data: 2011-2012)

Ratios	Mean	St. Dev.	Min	Median	Max
CT	29.462	76.884	0.606	11.127	1,285.099
CURRENT	2.203	1.356	0.147	1.900	19.227
DMKT	0.371	0.689	0.000	0.215	19.767
FLR	2.749	23.136	−1,197.462	2.149	731.265
NPM	7.598	46.278	−2,690.769	6.889	266.667
P/B	3.076	10.372	−150.678	2.131	666.267
P/E	24.355	60.550	0.340	16.270	2,000.000
ROA	7.194	5.337	−0.041	6.149	72.242
ROI	11.388	14.242	−0.059	9.441	752.146

Table J.4: Industry Summary Statistics (Yearly Data: 2002-2012)

Industry	Ratios	Mean	St. Dev.	Min	Median	Max
Energy	BVPS	18.226	25.379	−0.607	12.250	296.055
	CFS	3.852	5.605	−99.561	2.971	40.996
	CURRENT	1.884	1.321	0.199	1.526	16.182
	DY	1.302	3.178	0.000	0.262	34.716
	EPS	1.816	4.885	−111.750	1.470	31.500
Industrials	BVPS	15.273	12.867	−60.314	12.337	158.059
	CFS	3.062	3.500	−45.510	2.436	28.778
	CURRENT	2.046	1.090	0.119	1.817	10.909
	DY	1.488	2.627	0.000	1.023	47.221
	EPS	2.028	2.177	−16.940	1.640	21.050
	ROA	4.787	6.942	−123.594	5.259	32.030
Information Technology	CT	13.405	23.550	0.130	5.573	195.297
	CURRENT	2.986	3.148	0.172	2.130	35.698
	DMKT	0.244	0.746	0.000	0.122	16.850
	NPM	1.070	39.293	−844.101	4.165	88.383
	ROI	3.184	35.454	−743.224	7.187	69.625

APPENDIX K

MNL Results

Table K.1: MNL Results (Yearly Data: 2011-2012)

	Good	Poor
Intercept	0.929**	1.174***
	(0.304)	(0.242)
CT	−0.001*	−0.001**
	(0.001)	(0.000)
CURRENT	−0.595***	0.004
	(0.105)	(0.053)
DMKT	1.002**	1.155**
	(0.347)	(0.338)
FLR	0.139*	0.139*
	(0.062)	(0.062)
NPM	0.059***	0.001
	(0.01)	(0.002)
P/B	-0.01	−0.011′
	(0.006)	(0.006)
P/E	0.001	-0.001
	(0.001)	(0.001)
ROA	−0.129*	0.017
	(0.055)	(0.045)
ROI	-0.038	−0.081**
	(0.033)	(0.03)
Size	-0.051	-0.086
	(0.089)	(0.072)
Cox and Snell R-sq.	0.215	
Nagelkerke R-sq.	0.256	
McFadden R-sq.	0.132	
Likelihood-chi2(20)	330.317	
p	0.000	

The reference category is “superior”
Standard deviations in parentheses
*** p<0.001, ** p<0.01, * p<0.05, ’ p<0.1

Table K.2: MNL Results (Yearly Data: 2005-2012)

	Good	Poor
Intercept	0.28*	0.095
	(0.130)	(0.105)
CT	−0.001*	−0.001***
	(0.000)	(0.000)
CURRENT	−0.431***	0.033
	(0.046)	(0.022)
DMKT	1.974***	1.956***
	(0.18)	(0.179)
FLR	0.116***	0.117***
	(0.030)	(0.030)
NPM	0.002***	0.001***
	(0.000)	(0.000)
P/B	−0.014***	−0.02***
	(0.003)	(0.006)
P/E	0.000	0.000
	(0.000)	(0.000)
ROA	−0.092***	−0.061**
	(0.019)	(0.018)
ROI	0.005	0.000
	(0.011)	(0.011)
Size	-0.022	-0.064
	(0.041)	(0.037)
Cox and Snell R-sq.	0.196	
Nagelkerke R-sq.	0.223	
McFadden R-sq.	0.104	
Likelihood-chi2(20)	1091.863	
p	0.000	

The reference category is "superior"
Standard deviations in parentheses
*** p<0.001, ** p<0.01, * p<0.05, ' p<0.1

Table K.3: MNL Results (Quarterly Data: 2011-2012)

	Good	Poor
Intercept	0.928***	-0.039
	(0.174)	(0.139)
CT	−0.003***	−0.003***
	(0.001)	(0.001)
CURRENT	−0.340***	-0.01
	(0.050)	(0.030)
DMKT	2.596***	2.738***
	(0.207)	(0.204)
FLR	0.002	0.040***
	(0.002)	(0.007)
NPM	0.035***	0.001
	(0.005)	(0.001)
P/B	-0.004	−0.229***
	(0.010)	(0.026)
P/E	-0.001	-0.001
	(0.001)	(0.001)
ROA	−0.321***	-0.028
	(0.029)	(0.024)
ROI	0.047**	0.015
	(0.015)	(0.016)
Size	-0.051	−0.194***
	(0.047)	(0.046)
Cox and Snell R-sq.	0.275	
Nagelkerke R-sq.	0.311	
McFadden R-sq.	0.149	
Likelihood-chi2(20)	1303.906	
p	0.000	

The reference category is "superior"
Standard deviations in parentheses
*** p<0.001, ** p<0.01, * p<0.05, ’ p<0.1

Table K.4: MNL Results: Energy

	Good	Poor
Intercept	0.808**	0.637**
	(0.257)	(0.213)
BVPS	0.066***	0.062***
	(0.014)	(0.014)
CFS	0.127'	0.162*
	(0.072)	(0.071)
CURRENT	−0.635***	−0.226**
	(0.112)	(0.073)
DY	0.201***	0.153***
	(0.044)	(0.043)
EPS	−0.599***	−0.633***
	(0.101)	(0.099)
Cox and Snell R-sq.	0.176	
Nagelkerke R-sq.	0.200	
McFadden R-sq.	0.091	
Likelihood-chi2(20)	138.534	
p	0.000	

The reference category is “superior”
Standard deviations in parentheses
*** p<0.001, ** p<0.01, * p<0.05, ’ p<0.1

Table K.5: MNL Results: Industrials

	Good	Poor
Intercept	1.272***	0.950***
	(0.231)	(0.185)
BVPS	-0.004	-0.011
	(0.011)	(0.011)
CFS	0.344***	0.286***
	(0.061)	(0.059)
CURRENT	−0.542***	-0.052
	(0.096)	(0.058)
DY	0.123**	0.102**
	(0.040)	(0.038)
EPS	−0.267**	−0.171*
	(0.084)	(0.079)
ROA	−0.185***	−0.157***
	(0.020)	(0.019)
Cox and Snell R-sq.	0.157	
Nagelkerke R-sq.	0.177	
McFadden R-sq.	0.080	
Likelihood-chi2(12)	241.907	
p	0.000	

The reference category is "superior"
Standard deviations in parentheses
*** p<0.001, ** p<0.01, * p<0.05, ' p<0.1

Table K.6: MNL Results: Information Technology

	Good	Poor
Intercept	1.338**	0.580*
	(0.509)	(0.242)
CT	−0.051**	−0.016**
	(0.015)	(0.005)
CURRENT	−0.757***	0.044
	(0.190)	(0.046)
DMKT	2.272**	2.207**
	(0.656)	(0.649)
NPM	−0.037′	-0.015
	(0.019)	(0.016)
ROI	-0.014	-0.021
	(0.016)	(0.014)
Cox and Snell R-sq.	0.186	
Nagelkerke R-sq.	0.218	
McFadden R-sq.	0.107	
Likelihood-chi2(12)	95.108	
p	0.000	

The reference category is "superior"
Standard deviations in parentheses
*** p<0.001, ** p<0.01, * p<0.05, ' p<0.1

APPENDIX L

Likelihood Ratio Tests

Table L.1: Likelihood Ratio Tests (Yearly Data)

	P-value	
Ratios	2011-2012	2005-2012
Intercept	0.000	0.061
CT	0.016	0.001
CURRENT	0.000	0.000
DMKT	0.000	0.000
FLR	0.012	0.000
NPM	0.000	0.000
P/B	0.178	0.001
P/E	0.052	0.412
ROA	0.012	0.000
ROI	0.000	0.406
Size	0.475	0.205

Table L.2: Likelihood Ratio Tests (Quarterly Data: 2011-2012)

Ratio	P-value
CT	0.000
CURRENT	0.000
DMKT	0.000
FLR	0.000
NPM	0.000
P/B	0.000
P/E	0.446
ROA	0.000
ROI	0.000
Size	0.000

Table L.3: Likelihood Ratio Tests (Industry Data)

	P-value		
Ratios	Energy	Industrials	Information
Intercept	0.002	0.000	0.011
BVPS	0.000	0.456	×
CT	0.061	×	0.000
CURRENT	0.000	0.000	0.000
DMKT	×	×	0.000
DY	0.000	0.002	×
EPS	0.000	0.005	×
NPM	×	×	0.142
ROA	×	0.000	×
ROI	×	×	0.260
Size	×	×	×

APPENDIX M

Benchmark Accuracy Rates

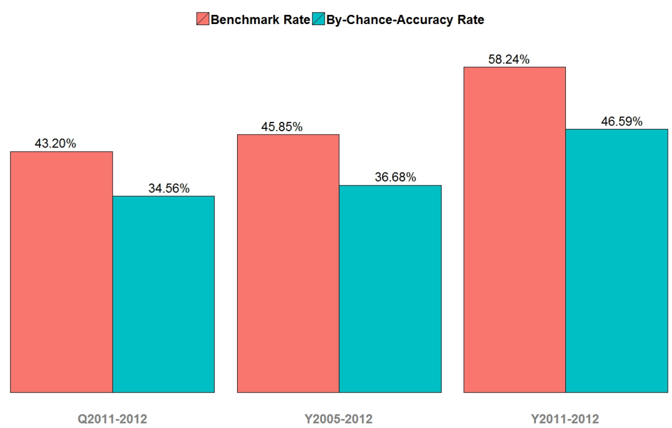

Figure M.1: Benchmark Accuracy Rates

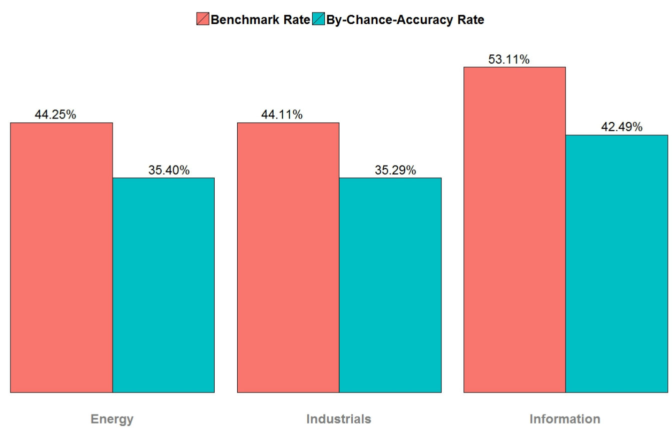

Figure M.2: Benchmark Accuracy Rates: Industry Level

APPENDIX N

Goodness-of-Fit Tests

Table N.1: Pearson and Deviance Statistics (Yearly Data)

		Chi-square	df	P-value
2011-2012	Pearson	5152.686	2710	0.000
	Deviance	2165.200	2710	1.000
2005-2012	Pearson	249991608.552	9978	0.000
	Deviance	9420.554	9978	1.000

Table N.2: Pearson and Deviance Statistics (Quarterly Data)

		Chi-square	df	P-value
2011-2012	Pearson	33771069.707	8074	0.000
	Deviance	7433.548	8074	1.000

Table N.3: Pearson and Deviance Statistics (Industry Data)

		Chi-square	df	P-value
Energy	Pearson	1429.163	1416	0.398
	Deviance	1388.085	1416	0.697
Industrials	Pearson	2953.789	2828	0.049
	Deviance	2799.204	2828	0.646
Information	Pearson	810.109	914	0.994
Technology	Deviance	792.778	914	0.998

APPENDIX O

Four-Layered Network

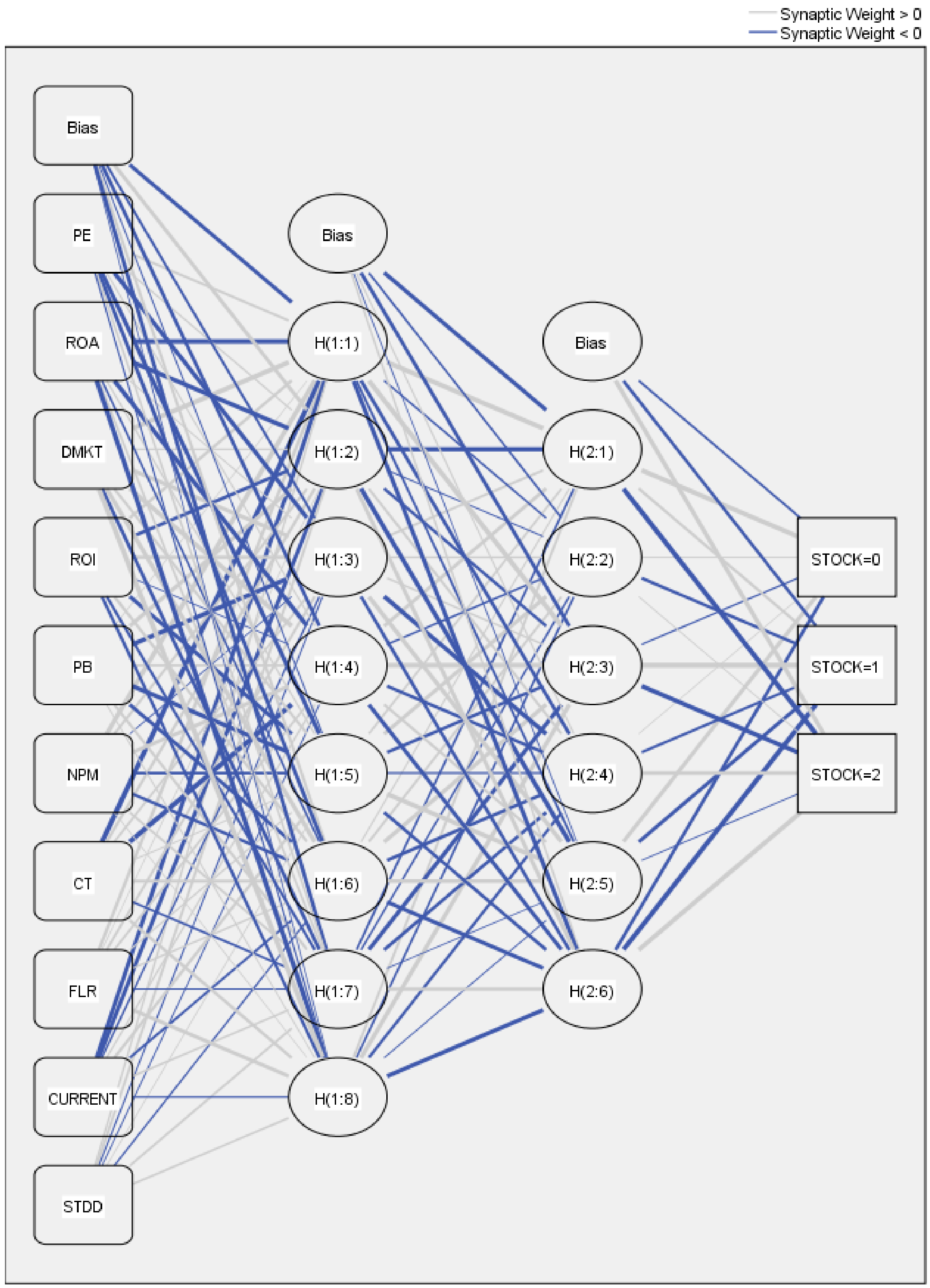

Figure O.1: Four-Layered Network

BIBLIOGRAPHY

BIBLIOGRAPHY

Agarwal, V. and R. J. Taffler (2007). Twenty-five years of the taffler z-score model: Does it really have predictive ability? *Accounting and Business Research 37*(4), 285–300.

Ahn, H., J. J. Ahn, K. J. Oh, and D. Kim (2011). Facilitating cross-selling in a mobile telecom market to develop customer classification model based on hybrid data mining techniques. *Expert Systems with Applications 38*(5), 5005–5012.

Akura, M. T. and K. Srinivasan (2005). Research note: Customer intimacy and cross-selling strategy. *Management Science 51*(6), 1007–1012.

Altman, E. I. (Sep., 1968). Financial ratios, discriminant analysis and the prediction of corporate bankruptcy. *The Journal of Finance 23*(4), 589–609.

Bichsel, F. (1967). Experience rating in subsets of risks. *ASTIN Bulletin 4*(3), 210–227.

Bridle, J. (1990). *Probabilistic interpretation of feedforward classification network outputs, with relationships to statistical pattern recognition.* in Neuro-Computing: Algorithms, Architectures and Applications. New York: Springer.

Bühlmann, H. (1967). Experience rating and credibility. *ASTIN Bulletin 4*(3), 199–207.

Bühlmann, H. and A. Gisler (2005). *A Course in Credibility Theory and its Applications.* Berlin: Springer Verlag.

Bühlmann, H., A. Gisler, and D. Kolloffel (2003). Multidimensional credibility applied to estimating the frequency of big claims.

Bühlmann, H. and E. Straub (1970). Glaubwürdigkeit für schadensatze. *Bulletin of Swiss Association of Actuaries 70*(1), 111–133.

Burgess, A. N. (1995). Loan risk analysis using neural networks. *The Journal of Targeting, Measurement and Analysis for Marketing 4*, 24–37.

Cameron, A. C. and P. Johansson (1997). Count data regression using series expansion: With applications. *Journal of Applied Econometrics 12*, 203–223.

Cameron, A. C. and P. K. Trivedi (1998). *Regression Analysis of Count Data*. Cambridge, U.K.: Cambridge University Press.

Chen, K. H. and T. A. Shimerda (1981). An empirical analysis of useful financial ratios. *Financial Management 10*, 51–60.

Cochrane, J. H. (1997). Where is the market going? uncertain facts and novel theories. *Economic Perspectives (Federal Reserve Bank of Chicago) 21*, 3–37.

Desjardins, D., G. Dionne, and J. Pinquet (2001). Experience rating schemes for fleets of vehicles. *ASTIN Bulletin 31*, 81–105.

DeVylder, F. (1978). Parameter estimation in credibility theory. *ASTIN Bulletin 1*, 99–112.

Dubey, A. and A. Gisler (1981). On parameter estimators in credibility. *Bulletin of Swiss Association of Actuaries*, 187–212.

Englund, M., M. Guillen, J. Gustafsson, L. H. Nielsen, and J. P. Nielsen (2008). Multivariate latent risk: A credibility approach. *ASTIN Bulletin 38(1)*, 137–146.

Englund, M., J. Gustafsson, J. P. Niesen, and F. Thuring (2009). Multidimensional credibility with time effects-an application to commercial business lines. *The Journal of Risk and Insurance 76(2)*, 443–453.

Greenwood, M. and G. Yule (1920). An inquiry into the nature of frequency distributions representative of multiple happenings with particular reference to the occurrence of multiple attacks of disease or of repeated accidents. *Journal of the Royal Statistical Society A83*, 255–279.

Hachemeister, C. A. (1975). *Credibility for regression models with an application to trend*. New York: Academic Press.

Hobarth, M. L. L. (2006). *Modeling the relationship between financial indicators and company performance-An empirical study for US listed companies*. Ph. D. thesis, Vienna University of Economics and Business Administration.

Jewell, W. S. (1973). Multidimensional credibility. *Operations Research Center*, 73–77.

Jewell, W. S. (1974). Exact multidimensional credibility. *Bulletin of Swiss Association of Actuaries 74*, 193–214.

Jones, F. (1987). Current techniques in bankruptcy prediction. *Journal of Accounting Literature 6*, 131–164.

Joseph Davis, R. A.-D. and C. J. Thomas (2012, October). Forecasting stock returns: What signals matter, and what do they say now? Technical report, The Vanguard Group Inc. (US).

Kaishev, V. K., J. P. Nielsen, and F. Thuring (2013). Optimal customer selection for cross-selling of financial services products. *Expert Systems with Applications 40*(5), 1748–1757.

Kamakura, W., M. Wedel, F. de Rosa, and J. Mazzon (2003). Cross-selling through database marketing: A mixed data factor analyzer for data augmentation and prediction. *International Journal of Research in Marketing 20*, 45–65.

Kanichukattu, K. J. and B. Abraham (2011). A count model based on mittag-leffler interarrival times. *Statistica 71*(4).

King, G. (1989). Variance specification in event count models: From restriction assumptions to generalizedd estimator. *American Journal of Political Science 33*, 762–784.

Larivière, B. and D. V. den Poel (2004). Investigating the role of product features in preventing customer churn, by using survival analysis and choice modeling: The case of financial services. *Expert Systems with Applications 27*(2), 277–285.

Larivière, B. and D. V. den Poel (2005). Predicting customer retention and profitability by using random forest and regression techniques. *Expert Systems with Applications 29*(2), 472–484.

Lewellen, J. (2004). Predicting returns with financial ratios. *Journal of Financial Economics 74*, 209–235.

McShane, B., M. Adrian, E. T. Bradlow, and P. S. Fader (2008, July). Count models based on weibull interarrival times. *Journal of Business and Economic Statistics 26*(3), 369–378.

Mowbray, A. H. (1914). How extensive a payroll exposure is necessary to give a dependable pure premium? *Proceedings of the Casualty Actuarial Society 1*, 24–30.

Norberg, R. (1982). On optimal parameter estimation in credibility. *Insurance: Mathematics and Economics 1*, 73–90.

Ohlson, J. A. (Spring, 1980). Financial ratios and the probabilistic prediction of bankruptcy. *Journal of Accounting Research, 18*(1), 109–131.

Pinquet, J. (1997). Allowance for cost of claim in bonus-malus systems. *ASTIN Bulletin 27*(1), 33–57.

Pinquet, J. (1998). Designing optimal bonus-malus system from different types of claims. *ASTIN Bulletin 28*(2), 205–220.

Pinquet, J. (2000). *Handbook of Insurance*, Chapter Experience rating through heterogeneous models, pp. 459–500. Kluwer Academic, Boston.

Refenes, A. N. (1995). *Neural Networks in the Capital Markets*. Chichester: Wiley Finance.

Seal, H. L. (1969). *Stochastic Theory of a Risk Business*. John Wiley and Sons, Inc., New York.

Sundt, B. (1979). A hierarchical credibility regression model. *Scandinavian Actuarial Journal*, 107–114.

Sundt, B. (1981). Recursive credibility estimation. *Scandinavian Actuarial Journal*, 3–21.

Taylor, P. (2004). The big picture on technical analysis. *Shareowner 17(3)*, 6–7.

Thuring, F. (2010). A credibility method for profitable cross-selling of insurance products. *Annals of Actuarial Science 6*(01).

Thuring, F., J. P. Nielsen, M. Guilln, and C. Bolanc (2012). Selecting prospects for cross-selling financial products using multivariate credibility. *Expert Systems with Applications 39*(10), 8809–8816.

Trippi, R. and E. Turban (1992). *Neural Networks in Finance and Investing*. New York: Probus Publishing.

Turk, A. F. (2006). The effect of financial ratios and market hype on short term stock prices. *Honors Projects Paper 30*, 96–103.

Upadhyay, A., G. Bandyopadhyay, and A. Dutta (2012). Forecasting stock performance in indian market using multinomial logistic regression. *Journal of Business Studies Quarterly 3*, 16–39.

Whitney, A. W. (1918). The theory of experience rating. *Proceedings of the Casualty Actuarial Society 4*, 274–292.

Winkelman, R. (1995). Duration dependence and dispersion in count-data models. *Journal of Business and Economic Statistics 13*, 467–474.

Zavgren, C. V. (1985). Assessing the vulnerability to failure of american industrial firms. *Journal of Business Finance & Accounting 12*(1), 19–45.

Zehnwirth, B. (1985). Linear filtering and recursive credibility estimation. *ASTIN Bulletin 15*(1), 19–38.

Zmijewski, M. E. (1984). Methodological issues related to the estimation of financial distress prediction models. *Journal of Accounting Research 22*, 59–82.