Essays on Failure Risk of Firms using Multivariate Frailty Models

A thesis submitted for the degree of Doctor of Philosophy

by

Francis Atsu

Department of Economics and Finance
College of Business, Arts and Social Sciences
Brunel University London
April 2016

Abstract

The post-2007 global financial crisis, characterised by huge firm losses, especially in the USA and Europe, initiated a new strand of literature, where default models are adjusted for unobserved risk factors, including measurement errors, missing firm specific and macroeconomic variables. These new models assume that default correlations are not only driven by observable firm-specific and macroeconomic factors, but also by unobserved risk factors. This thesis present three empirical essays.

The first essay estimates and predicts the within-sector failure rate and dependence of firms on the London Stock Exchange. The study offers an additive lognormal frailty model that accounts for both unobserved factors and regime changes. The analysis reveals that during distressed market periods the sector-based failure rates and dependencies tend to be high. The second essay proposes a novel approach based on a bias-corrected estimator to investigate the impact of informative firm censoring and unobserved factors on hazard rates of US firms. The approach uses inverse probability of censoring weighted scheme that explicitly accounts for firm specific factors, economic cycles, industry-level dependence and market activities induced by unobservable factors. The analysis shows that during distressed market periods the effect of informative censoring averagely increases the hazards rates, and varies across industries. The third essay employs a mixed effects Cox model to estimate the failure dependence caused by firms' exposure to country-based and group-level unobserved factors within the Eurozone. The empirical results show that a higher failure dependence among firms in groups of countries with similar economic and financial conditions than countries with different conditions.

Overall, the thesis contributes to the empirical literature on firm default in the broad area of corporate finance by offering a different approach of capturing default dependence and its variations during unfavourable market conditions and adjusting for the effects of non-default firm exit on active firms.

Dedicated to Jesus Christ and my family

Acknowledgements

First and foremost, great is God Almighty who is faithful and just; by His mercy and power that I was able to produce this piece of work. He offered me divine support and guidance throughout my studies in the UK. I say "God, take the glory, and let the blessings be mine".

Further, I am most grateful to my supervisor, Dr Mauro Costantini, for his consistent support, guidance and counsel while writing the thesis. He also supported me to land on a six month visiting researcher position at Cass Business school, UK. The experience gained there was of an immense help for my second essay. I sincerely express my profound appreciation to him.

Third, my special note of thanks go to Professor Ana-Maria Fuertes, Cass Business school, who offered me the opportunity to spend 6 months as a visiting researcher. I was inspired by her constructive feedbacks.

I am also indebted to some leading academics, who taught me some courses relevant to my area of research. Especially, the facilitators of the 2013 Oxford University Econometrics summer school, namely Professor Sir David F. Hendry, Dr Jennifer L. Castle and Dr Jurgen A. Doornik, Oxford University; Professor Siem Jan Koopman, University of Amsterdam; Professor Sebastian Laurent, Maastricht University. I benefited immensely from their outstanding research and teaching skills. Besides, Professor Dave Collett, Dr Alan Kimber, James Gallagher and Sandro Leidi introduced me to advanced topics in survival analysis which helped me in performing the empirical analysis of the second and third essays.

My sincere thanks and honour go to my beautiful wife, Rossetta Betty Prebi, for single-handedly taking care of our lovely children, Frances Etornam Atsu, Franca Delali Atsu, and Francis Selasi Atsu, over the course of my studies. Thanks to my children who have been praying to God to grant me success in all that I do.

My special thanks also go to Pastor Yemi Akinwande, who supported me spiritually and emotionally. He is not just a mentor, but also a friend and a father. Under his tutelage, I became a worker in the vineyard of Jesus Christ. May God grant him and his family long lives and sound health. I will like to thank my mother, Mary Akua Nambo, and my sisters, Francisca Atsufe and Mavis Nambo, for their continual support. May God bless them all.

A big thank you to Professor Lynn M. LoPucki (University of California, Los Angeles, School of Law) and Doug Irion, the database manager of the UCLA-LoPucki Bankruptcy Research Database, for providing me with free dataset of large US public companies.

Also, I would like to extend my profound gratitude to the Commonwealth Scholarship

Commission (CSC), UK, for selecting me as a commonwealth scholar. Specifically, CSC offered me a full financial support package and the opportunity to visit historical places (e.g. Buckingham Palace, Bank of England, Cumberland Lodge) across the UK.

Finally, I would to express my appreciation for the opportunity to be examined by Dr. Liang Han (University of Reading), and two members of the Department of Economics and Finance, Brunel University London, Dr. Nicola Spagnolo and Professor Menelaos Karanasos.

Declaration

A paper (Modelling corporate failure dependence of UK public listed firms) has been drawn from Chapter 2, with Dr. Mauro Costantini, and presented at: (i) 10th BMRC-DEMS Conference on Macro and Financial Economics/Econometrics Brunel University, 28-30th May 2014; (ii) School of Business and Economics, Faculty Seminar Presentation, Loughborough University, May 21, 2014; (iii) Economics Department Faculty Seminar Presentation, Reading University, May 28, 2014; (iv) Economics and Finance Faculty Seminar Presentation, Brunel University, March 19, 2014. Further, the paper will be submitted to the Journal of Corporate Finance.

In collaboration with Dr. Mauro Costantini and Professor Ana-Maria Fuertes, a paper, titled "Bias Correction in Hazards Rates: Evidence from USA Default Corporate Data", has been drawn from Chapter 3. The paper been presented at the 11th BMRC-DEMS Conference on Macro and Financial Economics/Econometrics Brunel University, 18-19th May 2015. In addition, the paper will be presented at the forthcoming INFINITI Conference on International Finance Trinity College Dublin, Ireland: 13-14 June 2016.

Dr Mauro Costantini and I presented material from Chapter 4, "Modelling Country and Group level Corporate Default Dependence: Evidence from Eurozone", at the Economics and Finance faculty seminar, Brunel University, November 25, 2015.

Table *of* Contents

List of Figures

List of Tables

Chapter 1

Introduction

The task of determining the financial health of businesses has attracted much attention among academics and practitioners over the last years. The seminal papers by Beaver (1967) and Altman (1968), who used univariate and multiple discriminant analyses (MDA, hereafter) respectively to discriminate financially healthy businesses from unhealthy ones, were the pioneers of a new way of estimating corporate failures using financial ratios. Since then, several models have been developed in the literature on corporate default. In the 1980s, some studies raised concerns about the basic assumptions of the MDA: independent variables normally distributed, probability of failure known a priori, and equal variance-covariance matrices across default and non-default categories (see, for example Zavgren, 1983; Karels and Prakash, 1987). To address these concerns, some authors proposed probit and logit models to estimate defaults without imposing those assumptions (see for example Ohlson, 1980; Zmijewski, 1984). These models have been then extended in order to deal with time-varying covariates and sector dynamics (see Gilbert et al., 1990; Platt and Platt, 1990; Charitou et al., 2004; Campbell et al., 2008).

However, when estimating failure rates, it is important to take into account the time to an event, e.g. failure or censoring. Therefore, the use of survival or duration models, where the dependent variables is the time a firm spends in an active group, has been encouraged. More specifically, duration models (e.g. Cox proportional hazard models) with static firm performance (firm characteristics are fixed over the period of existence; see Lane et al., 1986) and dynamic one (see Shumway, 2001; Kavvathas, 2001; Chava and Jarrow, 2004; Duffie et al., 2007; Bharath and Shumway, 2008), captured by time-varying specific and macroeconomic covariates, have been used.

In a related study on default rates in the presence of unobserved factors, Das et al. (2007) argued that those models based on the assumption that failure rates of firms are independent (firm failure rates are uncorrelated) are likely to produce bias in the estimates of hazard rates. Against this background, various works have taken into account failure dependence induced by unobserved risk factors at country or industry levels using frailty factors (see Duffie et al., 2009; Chava et al., 2011; Koopman et al., 2011; Koopman et al., 2012; Qi et al., 2014; Azizpour et al., 2015; among others), that are regarded as "a random component designed to account for variability due to unobserved individual level factors that is otherwise unaccounted for by the other predictors in the model" (Kleinbaum and Klein, 2012, pg. 326).

The thesis investigates the impact of unobserved risk factors on firms by using firm failure prediction models, and draws motivation from the argument that taking into account firm failure dependence results in more accurate estimates of hazard rates. As such, the study contributes to the empirical literature of corporate default in three regards. First, it offers a new way to measure default correlations in hazard rates during distressed market periods. Second, it shows how to capture the effects of non-default firm exit on active firms specifically during unfavourable conditions. Finally, it provides an approach to control for the effects of country-based and group level unobserved factors on firms. The results suggest that neglecting the effects of default correlations and non-default firm exit, and country and group levels unobserved effects may likely lead to underestimation of default rate of firms especially during distressed market periods.

The above contributions are offered within three empirical essays. The first essay, which contains the second chapter, explores the effects of unobserved risks factors under two regimes, namely normal and distressed market periods, on firms in the UK. The second essay, included in the third chapter, performs bias correction in hazard rate models that do not explicitly account for informative firm censoring using US corporate data. The third essay, contained in the fourth chapter, examines the effects of unobserved risk factors at country and group levels on firms listed in the Euro area.

Chapter 2 examines corporate failure dependence induced by firms' exposure to unobserved risk factors (frailty factors) under extreme market conditions on the London Stock Exchange over the period 1985-2012. These factors may include missing and measurement errors in covariates, variations in managerial flexibility, regulatory requirements, firm culture, cost control, and employee skills (see e.g. Lancaster, 1990; Hougaard, 2000; Chava et al., 2011). The chapter contributes to the extant literature in some respects. First, we propose an additive lognormal frailty model that accounts for both extreme and normal market regimes using a lognormal distribution. The lognormal distribution offers much more flexibility in estimating dependence among units (e.g. firms and countries) within a multivariate context (see Hougaard, 2000; Dutchateau and Janssen, 2008; Wienke, 2011; Hangal, 2011, among others) than other distributions, such as the Gamma one (see Chapter 2). As a result, the additive lognormal frailty model is likely to provide more accurate estimates of corporate failure dependence under extreme market conditions. Second, to the best of our knowledge, this is the first work that explores the corporate failure dependence at industry level in the UK using listed firms on the London Stock Exchange. Lastly, using a naïve recursive extraction approach, we compare the one-step ahead prediction performance of the additive lognormal model with that of the multiplicative gamma frailty model by Chava et al. (2011).

In our empirical analysis, we employ covariates used in previous studies (see Shumway, 2001; Duffie et al., 2007; Bharath and Shumway, 2008), such as one year trailing stock return, one year trailing market return, distance to default probability, 3 month T-bill rate, market value of equity (or equity), firm age, excess return, total assets, total liabilities to total assets, and net income to total assets.

The empirical analysis delivers three main results. First, the frailty factor is significant across all the specifications, a result consistent with findings in previous papers (see Duffie et al., 2009; Chava et al., 2011; Koopman et al., 2011, 2012; Qi et al., 2014; Azizpour et al., 2015; among others). This result shows evidence of firm failure clustering on the London Stock Exchange, especially during distressed periods. Second, the regime switch factor, which accounts for extra failure rate variations during distressed periods, is positive and significant for all the models. The factor increases as one moves from a less severe distressed period to a more severe distressed one. This adjustment in the failure rates during distressed periods is more accurate when using the additive lognormal frailty model than the multiplicative gamma frailty one. Lastly, the additive model tend to produce more accurate estimates and extracts (predicts) of the industry level frailties and dependence as compared to the multiplicative gamma frailty model. These results are supported by a better goodness-of-fit of our model. In order to deal with the variations of unobserved factors during distressed market periods, characterised by higher variations in hazard rates, we construct and use root mean square deviations metric to measure the deviations of these factors from their expected value. High values of this metric are desirable as opposed to low values for the classical root mean square error, which is not an appropriate measure when aiming to capture the effects of unobserved factors.

Chapter 3 aims to estimate the failure rates of public listed firms on the following exchanges NYSE, NASDAQ and NYSE MKT LLC, over the period 1980-2013, while considering the potential effects of informative firm censoring. Firms may leave the market voluntarily or involuntarily. This activity can reveal the financial conditions of such firms and, to some extent, of other firms. For instance, two financially healthy firms may form a synergy to enhance corporate market power, profitability, and shareholders' wealth through the activities of mergers and acquisitions (M&As). Also, a financially distressed firm may be willing to be acquired in order to survive (see Andrade et al., 2001; Lambrecht and Myers, 2007; Alexandridis et al., 2010, among others). In addition, firms with weak corporate governance structures are more likely to be delisted (see Marosi and Massoud, 2008; Hostak et al., 2013, among others). All these market activities are likely to impact on the hazard rates of the surviving firms, and neglecting to account for this may generate bias in the hazard rate estimates.

This study proposes a novel empirical approach for estimating failure rates conditional on informative firm censoring, unobserved factors and extreme market conditions. In particular, the study aims to estimate the probability of non-default firm exit and the corresponding impact on hazard rate. In doing so, the estimates of the hazard rate are compared to those obtained using the multivariate gamma frailty and additive lognormal frailty models of Chava et al. (2011) and Atsu and Costantini (2015), respectively. These two models do not explicitly account for informative censoring and are likely to underestimate hazard rates, especially during distress market periods.

This work contributes to corporate failure dependence literature by performing bias-correction in the models of Chava et al. (2011) and Atsu and Costantini (2015), and it applies inverse probability of censoring weighted (IPCW, hereafter) (see e.g. Robins, 1993; Robins and Finkelstein, 2000; Scharfstein and Robins, 2002). This allows us to quantify the potential effects of informative firm censoring, as dynamic weights, from two perspectives. First, we construct the weights by using time varying firm-specific and macroeconomic factors. Second, we combine firm-specific and macroeconomic factors with industry level unobserved factors to estimate the weights. This is done because some market activities, such as M&As, tend to cluster by industry (see e.g. Andrade et al., 2001; Harford, 2005).

As for the regression analysis, we employ the following covariates: one year trailing stock return, one year trailing market return, distance to default probability, firm age, 3 month T-bill rate, and industry level distress indicator. The analysis points to three main results. First, the distressed indicator is positive and significant in all the models, providing the evidence of higher hazard rates during distressed market periods. The bias-corrected models adjust hazard rates up during distressed periods more accurately than the models of Chava et al. (2011) and Atsu and Costantini (2015). Second, models by Chava et al. (2011) and Atsu and Costantini (2015) underestimate the effects of unobserved risk factors and failure rate dependence as compared to the weighted models. Lastly, the bias-corrected models are comparatively more efficient than the benchmark models, as they tend to produce smaller standard errors.

Chapter 4 investigates failure rates and dependence caused by firms' exposure to both country-based and group level unobserved risk factors in the Euro area over the period 1994Q1-2014Q4. The Euro area offers some advantages to its members, such as free movement of trade and capital, reduction of transaction costs, elimination of exchange rate uncertainty, price transparency, and potential development and integration of financial markets, but it also implies costs to state members, such as the loss of monetary independence. The recent financial crisis and the European banking and debt crises have hit the Euro countries in different ways. While the PIIGS economies (Portugal, Ireland, Italy, Greece, and Spain) have been hit harder, other countries have suffered less, even though the contagion effect have propagated to Belgium and France (see Metiu, 2012; Arghyrou and Kontonikas, 2012; Ludwing, 2014). As a result, financial conditions of firms have been seriously affected (see Bhattacharjee et al., 2009; Bonfim, 2009; Chen, 2010; Tang and Yan, 2010; Jacobson et al., 2013, among others), and businesses within the Euro area have struggled to survive, with a large impact on hazard rates. This suggests that internal (country-based) and external (group level) unobserved risk factors should be considered when estimating default rates.

The chapter contributes to the extant literature in some respects. First, this study takes into account the impact of external unobserved factors (due to the Euro membership) along with the internal ones, while previous studies treat countries as standalone entities. Second, this study is the first to estimate the default rate of firms and their dependence for 11 selected Eurozone countries, with a focus on the following groups of countries: the PIIGS, the PIIGSB (Belgium is part of the PIIGS), the PIIGSF (France belongs to the PIIGS), and the PIIGSBF (both Belgium and France are members of the PIIGS).

To estimate the default rates, this study uses a mixed effect Cox model, which allows us to nest frailty factors at country and group levels. We use one year trailing stock return, one year trailing market return, distance to default probability, firm age, and 3 month T-bill rate as covariates. The regression analysis offers three results. First, there is a significant evidence of failure dependence caused by firm's exposure to country level unobserved factors. Second, when countries are grouped together, failure clustering tend to be larger, as firms are subject to an extra risk due to the impact of unobserved factors at the group level. Third, models that do not account for the distance to default probability tend to perform poorly as compared with their counterparts.

Chapter 5 presents conclusions and recommendations.

Chapter 2

Modelling corporate failure dependence of UK public listed firms

2.1 Introduction

The estimates of failure probability and its correlation play a central role in contemporary risk management for corporations, regulators, investors and academics. In particular, they can be used by: a wide range of stakeholders to explore how economic cycles and corporate default risk are related over a period; rating agencies to rate firms; banks and bank regulators to determine minimum capital requirements; financial institutions to discriminate good credit applicants from the bad ones (Shumway 2001; Duffie et al, 2007, Duan et al., 2012, among others).

In this study, we explore the dynamics of corporate failure dependence and its variations across various sectors on the London Stock Exchange (LSE, henceforth) over the period 1985-2012. To this end, we use a multivariate frailty reduced form model that accounts for unobserved factors.

Literature broadly groups credit risk models into structural and reduced form models, given the role that information plays in modelling default risk (see e.g. Jarrow and Turnbull, 1992; Jarrow and Turnbull, 1995; Duffie and Singleton, 1999; Duffie and Lando, 2001; Jarrow and Protter; 2004; Giesecke, 2006; among others). However, reduced form approaches have received more attention than the structural ones (Jarrow, 2001, Jarrow and Protter, 2004; Duan et al. 2012; Dionne and Laajimi, 2012; Figlewski et al., 2012; Yeh et al., 2015), since these models are primarily based on the information available to the market.[1] In this chapter, we employ the reduced form approach, and to estimate the

[1] For a comprehensive comparison between structural and reduced form models, see Jarrow and Protter

parameters of the model and within-sector dependence, we consider covariates used in previous works, such as one year trailing market return, one year trailing stock return, 3 month T-bill rate, distance to default probability, excess return, net income to total assets, total liabilities to total assets, stock volatility, market value (or equity), and firm age (see Shumway, 2001; Duffie et al., 2007; Bharath and Shumway, 2008).

This chapter makes some contributions to the literature on corporate finance. First, we propose an additive lognormal frailty model with two regime changes (distressed and normal regimes). While the literature predominately features gamma distribution (see e.g. Chava et al., 2011; Wienke, 2011), we use the lognormal distribution as it offers much more flexibility in modelling the dependence structures within a multivariate context (see e.g. Hougaard, 2000; Duchateau and Janssen, 2008; Wienke, 2011). The lognormal distribution is positively skewed and the dependence measure (or association) is directly proportional to the skewness of the distribution: the higher the value of association, the greater the skewness which makes the right tail longer (Lee and Wang, 2003). As the data on corporate failure is highly skewed during distressed periods, a power transformation of the frailties as to make them normal-like may help to better capture the dependence on a log-scale (Hougaard, 2000). Therefore, under extreme market conditions, the lognormal tends to properly explain the frailties and the corresponding dependence structures, and the additive lognormal frailty model may provide more accurate information on corporate failure dependence as compared to models which use gamma distribution. Second, we investigate the dynamics of corporate failure dependence on the LSE. To the best of our knowledge, this is the first study to look at corporate failure dependence in the UK. We also test the robustness of our model under different levels of sector distress (degrees of departure from normal market conditions), given the fact that the effects of unobserved sector specific factors tends to be more pronounced when markets move to severe distressed conditions, and compare its performances with those of the model by Chava et al. (2011). The comparison is carried out using measures of goodness-of-fit. Lastly, and this is another novelty of the study, we investigate one-step ahead predictive performances of our model and the model by Chava et al. (2011), using a naïve recursive extraction approach.

Our empirical analysis delivers three main results. First, the frailty factor is always significant across all the model specifications as in previous studies (see Duffie et al., 2009; Chava et al., 2011; Koopman et al., 2011, 2012; Qi et al., 2014). Further, the significance of the within-sector frailties provides evidence of firm failure clustering, which tends to occur more during distressed market conditions. Second, the adjustment factor in hazard

(2004).

rate during distressed market periods is also significant, and this implies that firms on the LSE are more inclined to move faster towards failure. Lastly, the additive lognormal frailty model tends to better estimate and predict within-sector frailties and dependencies than the multiplicative gamma frailty model when moving away from normal market conditions, as it is evidenced by information criteria results. This seems to favour the use of the additive lognormal frailty model when estimating and predicting correlations and failure rates among firms during distressed market conditions in the UK.

The rest of the chapter is organized as follows. Section 2.2 reviews previous studies on the dynamics of corporate default risks. Section 2.3 presents methodology and data. Section 2.4 discusses the empirical findings and Section 2.5 concludes the study.

2.2 Literature review

Literature broadly classifies failure prediction models into first, second and third generation models. The first generation models (FGMs) primarily employed discriminant analysis (see e.g. Beaver, 1967, 1968; Altman, 1968; Deakin, 1972) and risk indexing (e.g. Tamari, 1966; Moses and Liao, 1987) to compute credit scores as to differentiate healthy firms from financially distressed ones. The second generation models (SGMs) (see Ohlson, 1980; Zmijewski, 1984; Zavgren, 1985; Gentry, 1987) improve over the FGMs by using binary response models, such as logit and probit, where the dependent variable of failed and active firms is assigned value 1 and 0, respectively. The third generation models (TGMs), which largely feature duration analysis (Shumway, 2001; Kavvathas, 2001; Chava and Jarrow, 2004; Hillegeist et al., 2004; Duffie et al., 2007; Bharath and Shumway, 2008; among others), account for changes in firms' characteristics over a period, as opposed to FMGs which consider only one-period data set of firms. In addition, these models accounts for time in different ways using: (i) time varying covariates; (ii) time spent by a firm in the surviving group as the dependent variable; and (iii) firm age as a significant prognostic factor (see Shumway, 2001).

In their study on US corporate default data, Das et al. (2007) observed excess default correlation induced by unobserved factors (or frailty factors), and showed that models based on the assumption that corporate defaults are conditional independent after adjusting for observable factors tend to underestimate default clustering. To address this issue, Duffie et al. (2009) developed a dynamic frailty model, which explicitly accounts for default correlations, to estimate default rate of a set of 2793 US non-financial firms for monthly data over the period 1979-2004. The frailty factor is allowed to revert to its mean in the event of shocks. The empirical analysis confirmed the existence of common latent factors (frailty), that account for about 40 percent extra variations of corporate default rates, even after taking observable factors into account, with very high firm default prognostic effects. Using various goodness-of-fit and quantile tests, the authors concluded that corporate default models that do not incorporate unobserved factors produce downward biased estimates. The use of such models may produce misleading results concerning the minimum capital requirement of firms, which are likely to hold capital that does not necessarily reflect their risk profiles.

Using an augmented dataset of Duffie et al. (2009), Koopman et al. (2011) proposed a new non-Gaussian panel data model that incorporates the principal components of a large data set of macroeconomic and financial covariates of US firms. The authors argued that their framework is more appropriate for the estimation and forecast of the dynamic corporate failure rates using both observable and unobservable risk factors. The empirical results showed that the dynamic frailty factor plays a crucial role, even after controlling for at least 80 percent of the changes in over 100 macroeconomic and financial covariates, and models with frailty outperform those without frailty in out-of-sample analysis. Koopman et al. (2011) argued that including frailty factors in corporate default models enhances the estimation and forecasting abilities of such models.

In a related study, Koopman et al. (2012) proposed a new decomposition approach for systematic default risk. Using high-dimensional, nonlinear and non-Gaussian dynamic factor models, the work simultaneously measured the effect of macroeconomic/financial, frailty and industry-level risk factors on US corporate default rate variations (or clustering) over the period 1971Q1-2009Q1. The empirical results revealed that: (i) systematic and industry factors accounts for approximately 35% of default rate variations, and about 33% of the default clustering is accounted for by the macroeconomic and financial factors; (ii) the frailty factor captures about 40% of the default rate variation; and (iii) 25% of the latter is accounted for by the industry effects. The authors argued that the frailty component plays a major role as compared to other components because: (i) it tends to capture a higher portion of corporate default rate variations before and during times of crisis, as their unified technique is able to detect systematic credit risk build-up in the years 2002-2008; and (ii) it accounts for missing sources of default rate variations. Koopman et al. (2012) concluded that the dynamics of default rate variations are more appropriately captured when frailty effects are taken into account.

Qi et al. (2014) tested the significance of an unobserved systematic risk factor in a corporate default prediction exercise using a univariate frailty approach by Duffie et al. (2009) for the US corporate data of 3650 active and 508 failed firms over the period 1979-2010. The results showed that the unobserved risk factor is more highly informative than the observed factors in the in-sample analysis, while its predictive power improvement is very lower for the out-of-sample forecasting exercise.

Using financial and industrial default timings in the US over the period 1970-2012, Azizpour et al. (2015) examined whether corporate default clustering in US are caused by frailty and contagion. The authors developed a new model and various tests using time dependent observable factors, dynamic frailty factor and past defaults. The results showed that frailty and contagion induce default clustering after accounting for the effects of macroeconomic and firm specific factors, and in addition the past firm failures tend to explain a portion of the conditional default rates.

Chava et al. (2009) argued that economy-wide frailty models do not provide relevant industry-level information about default rates, since they do not account for changes in specific industry-level. Therefore, the authors developed a multiplicative frailty model to estimate and predict firm defaults and recovery rates using observable and unobservable covariates for the US data over the period 1980-2008. The estimates are used to model and predict the loss distribution of bonds and loans. Chava et al. (2011) improved over the model by Duffie et al.(2009) by adopting two regimes (normal and distressed industry level periods) frailty modelling framework. The choice of a shared frailty allows the authors to explore the contagion effects at the industry level. The model of Chava et al. (2011) assigns a different frailty factor to each industry, whereas the model by Duffie et al.'s (2009) uses a single frailty for all industries. The empirical results revealed that default models that control for regime switching and industry level frailties have a higher explanatory power of portfolio loss dynamics as compared to economy-wide frailty models.

2.3 Methodology and data

In this section, we first present our additive lognormal frailty model and the multiplicative gamma frailty of Chava et al. (2011), and then we describe the data.

2.3.1 Additive lognormal frailty model

Our additive lognormal frailty model is based on the approach of Clayton (1978). Let $T \in [0, \infty)$ be the time to event or time until a firm either fails or leaves the sample as a result of non-failure events (e.g. mergers and acquisitions). Our data set contains s clusters (sectors) and in each cluster there are n_i members (firms) (see Duchateau and Janssen, 2008). In our sample, the sum of firms across all the sectors is the total number of firms, $n = \sum_{i=1}^{s} n_i$. Given a time horizon $[0, T^*]$, staggered firm entry is allowed and some firms may leave the sample period due to non-failure events. In addition, some firms may experience failure event or survive beyond the end of the sample period, T^*, and a firm is considered censored if it leaves the sample period through non-failure reasons or survives beyond T^*. The information consists of the set $(T_{ij}, \delta_{ij}, X_{ij}(t), \tilde{u}_i)$ for $i = 1, ..., s$ and $j = 1, ..., n_i$. The term T_{ij} is the event time (either failure or censored time) of the jth firm in the ith sector, δ_{ij} is the corresponding censoring indicator which takes value 1 when T_{ij} is the failure time and 0 if T_{ij} is the censoring time, and $\delta_i = \sum_{j=1}^{n_i} \delta_{ij}$ is the total number of failures in the ith sector. The vector $X_{ij}(t)$ is the set of time-varying covariates for the jth firm in the ith sector in the counting process style of input. Finally, $\tilde{u}_i$ is the unobserved information or the frailty term for ith sector. The frailty factor is defined as "a random component designed to account for variability due to unobserved individual-level factors that is otherwise unaccounted for by the other predictors in the model" (Kleinbaum and Klein, 2012, page 326). These factors may include missing and measurement errors in covariates, variations in managerial flexibility, regulatory requirements, firm culture, cost control, and employee skills (see e.g. Lancaster, 1990; Hougaard, 2000; Chava et al., 2011).

We use the classical shared frailty modelling approach of Clayton (1978) to derive our additive lognormal frailty model. The classical shared frailty model is based on the Cox proportional hazard (PH) semi-parametric framework and is defined as follows:

$$h_{ij}(t) = h_0(t)\tilde{u}_i exp(X_{ij}(t)\beta), \tag{2.1}$$

where $h_{ij}(t)$ is the conditional hazard rate for the jth firm in the ith sector (conditional on the frailty factor, $\tilde{u}_i$), $h_0(t)$ is an arbitrary baseline hazard and β is a p-dimensional vector of coefficients of the covariates, $X_{ij}(t)$. We rewrite the frailty factor $\tilde{u}_i$ in terms of a random effect or log-frailty as: $\tilde{w}_i = \log \tilde{u}_i$ or $\tilde{u}_i = exp(\tilde{w}_i)$. Then, equation (2.1) becomes:

$$\begin{aligned} h_{ij}(t) &= h_0(t) \exp(log(\tilde{u}_i))exp(X_{ij}(t)\beta) \\ &= h_0(t)exp(X_{ij}(t)\beta + \tilde{w}_i). \end{aligned} \tag{2.2}$$

Equation (2.2) represents the classical lognormal shared frailty model. It contains two terms: the fixed effects term, which involves the covariates, and the random term, $\tilde{w}_i$, with an expected value, $E(\tilde{W}) = 0$ and a finite variance, $Var(\tilde{W}) = \gamma$. We follow Chava et al. (2011) to construct the log-frailty term as a combination of sector-specific log-frailty term, w_i, and a time-varying sector distress indicator, $Z_i(t)$, which takes value 1 for distressed sectors at time t and 0 otherwise. As such, we have:

$$\begin{aligned}\tilde{w}_i(t) &= \log \tilde{u}_i = \log(u_i \Delta^{Z_i(t)}) \\ &= \log u_i + \log \Delta Z_i(t). \end{aligned} \tag{2.3}$$

Equation (2.3) can be re-written as:

$$\tilde{w}_i(t) = \pi Z_i(t) + w_i, \tag{2.4}$$

where $\pi = \log(\Delta)$ is the additive factor in the regime-switch lognormal frailty context that accounts for the extra variations in hazard rates induced by distressed market periods.

Substituting equation (2.4) into (2.2), the additive lognormal frailty model (regime-switch lognormal frailty model) is given by:

$$h_{ij}(t) = \begin{cases} h_0(t)exp(X_{ij}(t)\beta + \pi Z_i(t) + w_i) & \text{if sector } i \text{ is distressed,} \\ h_0(t)exp(X_{ij}(t)\beta + w_i) & \text{otherwise,} \end{cases} \tag{2.5}$$

where $h_0(t)exp(X_{ij}(t)\beta + w_i)$, and $h_{ij}(t) = h_0(t)exp(X_{ij}(t)\beta + \pi Z_i(t) + w_i)$ are the hazard functions for normal and distressed periods, respectively.

The classical shared lognormal frailty model is a special case of our additive lognormal frailty model when $\pi = 0$. The shared lognormal frailty model does not incorporate regime changes in the impact of the lognormal frailties. Although the multiplicative gamma frailty model may show high predictive power (see Chava et al., 2011), we argue that our additive lognormal frailty model offers much more flexibility than the gamma frailty model due to the properties of the lognormal distribution within the multivariate context (Hougaard, 2000; Duchateau and Janssen, 2008; Wienke, 2011). This flexibility stems from the dependence between the right tail of the distribution and the association parameter (Lee and Wang, 2003), and its power transformation property (Hougaard, 2000).

To estimate the parameters in equation (2.5), we use the penalised partial likelihood (PPL, hereafter) approach of McGilchrist and Aisbett (1991):

$$l_p(\beta, \pi, \gamma | w) = l_{part}(\beta, \pi | w) - l_{pen}(\gamma | w), \tag{2.6}$$

where

$$l_{part}(\beta, \pi | w) = \sum_{i=1}^{s} \sum_{j=1}^{n_i} \delta_{ij} \left(X_{ij}(t)\beta + \pi Z_i(t) + w_i - \log \left(\sum_{j \in R(T_{ij})} exp(X_{ij}(t)\beta + \pi Z_i(t) + w_j) \right) \right), \tag{2.7}$$

which is the conditional likelihood given the log-frailties and

$$l_{pen}(\gamma|w) = \frac{1}{2\gamma}\sum_{i=1}^{s} w_i^2, \tag{2.8}$$

represents the penalised term (the distribution of the log-frailties). This term penalises the likelihood by subtracting large values of the penalty term from the full data log likelihood, if the real values of the log frailties are far from their mean (see Duchateau and Janssen, 2008). The term $R(T_{ij})$ in equation (2.7) is the risk set (the set of surviving firms or firms still at the risk of an event). The PPL does not depend on the baseline hazard function, making it possible to estimate the parameters of the likelihood without knowing the shape of the baseline hazard rate. This characteristic of PPL makes our estimates robust, regardless of the shape of the baseline hazard rate (see e.g. Cox, 1975; Duchateau and Janssen, 2008; Allison, 2010), although estimates can be, to some extent, not fully efficient, but this inefficiency is normally immaterial (see Efron, 1977). However, the estimates are consistent and asymptotically normal (see e.g. Cox, 1975; Allison, 2010).

Let $\beta^* = (\beta, \pi)$ be the coefficients of the following covariates $X = (X_{ij}(t), Z(T_{ij}))$. Equation (2.7) can be the re-written as follows:

$$l_{part}(\beta^*|w) = \sum_{i=1}^{s}\sum_{j=1}^{n_i}\delta_{ij}\left(X\beta^* + w_i - \log\left(\sum_{j\in R(T_{ij})} exp(X\beta^* + w_j)\right)\right). \tag{2.9}$$

Therefore equation (2.6) becomes:

$$l_p(\beta^*, \gamma|w) = \sum_{i=1}^{s}\sum_{j=1}^{n_i}\delta_{ij}\left(X\beta^* + w_i - \log\left(\sum_{j\in R(T_{ij})} exp(X\beta^* + w_j)\right)\right) - \frac{1}{2\gamma}\sum_{i=1}^{s} w_i^2. \tag{2.10}$$

For any value of the log-frailty variance, γ, we employ the marginal log-likelihood in Ripatti and Palmgren (2000) (see also Therneau and Grambsch, 2000; Therneau et al., 2003; SAS/STAT 13.2) to derive the extended PPL of equation (2.10):

$$l_m(\beta^*, \gamma) = -\frac{1}{2}\log(\gamma I) + \log(\int exp[l_p(\beta^*, \gamma)]dw), \tag{2.11}$$

where I is the identity matrix of order $s \times s$ and s is the number of sectors in the sample. We use the approximation of Ripatti and Palmgren (2000) to derive the likelihood in equation (2.11):

$$l_m(\beta^*, \gamma) \approx -\frac{1}{2}\log(\gamma I) + \log(|H_{22}(\beta^*, \gamma, w^*)|) - l_p(\beta^*, \gamma, w^*), \tag{2.12}$$

where H is the negative Hessian of the PPL for a given value of γ. We use the $PHREG$ procedure in SAS to maximize the likelihood in equation (2.12). For the variance of the frailty, $\tilde{u}_{it}$, which follows a lognormal distribution with an expected value, $E(\tilde{U}) = 1$, and a finite variance, $Var(\tilde{U}) = \theta$, it is required that $\gamma = \log(\theta+1)$ (see Duchateau and Janssen, 2008).

In our empirical analysis, we compare the performance of the first specification of our model with the multiplicative frailty model (see Chava et al. 2011) in order to empirically ascertain whether the additive lognormal frailty model is comparatively better than the latter. In what follows, we briefly describe the multiplicative gamma frailty model of Chava et al. (2011):[2]

$$h_{ij}(t) = \begin{cases} h_0(t)u_i\Delta^{Z_i(t)}exp(X_{ij}(t)\beta) & \text{if sector } i \text{ is distressed,} \\ h_0(t)u_iexp(X_{ij}(t)\beta) & \text{otherwise.} \end{cases} \tag{2.13}$$

For estimation feasibility, the authors assumed that the sector frailties follow a two parameter gamma distribution, i.e. $u_i(t) = G(A_i(t), C_i(t))$ with the shape parameter $A_i(t) = 1/\theta(t) + \sum_{j=1,T_{ij}<t}^{n_{ij}} \delta_{ij}$ and scale parameter $C_i(t) = 1/\theta(t) + \sum_{j=1,T_{ij}<t}^{n_{ij}} H(T_{ij})$, where $H(T_{ij}) = \int_0^{T_{ij}}(\Delta^{Z_i(t)}exp(X_{ij}(t)\beta)dt$. Based on the above assumption, the authors derived the sample marginal likelihood for all sectors as

$$l(\theta, \Delta, \beta) = \sum_{i=1}^{s} l_i(\theta, \Delta, \beta), \tag{2.14}$$

where

$$\begin{aligned} l_i(\theta, \Delta, \beta) = \log\Gamma(\delta_i + 1/\theta) - \log\Gamma(1/\theta) - (1/\theta)\log(\theta) + \sum_{j=1}^{n_i}\delta_{ij}(X_{ij}(T_{ij})\beta \\ +Z_i(T_{ij})\log(\Delta)) - (\delta_i + 1/\theta)\log(1/\theta + \sum_{j=1}^{n_i} H(T_{ij})), \end{aligned} \tag{2.15}$$

for each sector i. The term $\Gamma(.)$ is the gamma function with an expected value of 1 and a finite variance, θ.[3]

In order to select the best specification of our model and compare it with that of Chava

[2]For further details, readers can refer to Chava et al. (2011). Here, we change some of the notations in Chava et al. (2011) to ease the comparison of the two models.

[3]Chava et al. (2011) applied the expectation maximization (EM) approach, while we use PPL technique. However, for a gamma distribution, the EM and PPL procedures lead to the same results (for details in this respect, see Duchateau and Janssen, 2008). For consistency, therefore, we maximise equation (2.14) using the PPL procedure.

et al. (2011), we use the Akaike Information Criterion (AIC), the corrected Akaike Information Criterion ($AICC$), and the Bayesian Information Criterion (BIC) as defined below:

$$AIC = -2logL + 2k, \tag{2.16}$$

$$AICC = AIC + \frac{2k(k+1)}{n-k-1}, \tag{2.17}$$

$$BIC = -2logL + klogn, \tag{2.18}$$

where $-2LogL$ is the partial likelihood which is obtained by using the rank of events (Singer and Willett, 2003), k and n denote the number of parameters and events, respectively (see Xie, 1994; Raftery, 1995).

2.3.2 Data

Data are taken from DataStream and Worldscope for the London Stock Exchange (LSE). It covers the period 1985-2012 due to data availability. Our sample contains 889 firms, which consists of 524 active, 174 merged or acquired and 191 failed firms, which translates into 13,343 yearly firm observations. To study the within-sector dependencies and frailties, we employ 29 subsectors from the 10 major DataStream sectors on the LSE (see Table 2.1). As regards the definition of failure, we follow the convention of legal definition of failure (see, e.g., Charitou et al., 2004; Christidis and Gregory 2010; Tinoco and Wilson, 2013) and select firms in this category. Given our sample, we specifically employ the UK insolvency Act 1986 to select failed firms. The Act states that, “A company is insolvent (unable to pay its debts) if it either does not have enough assets to cover its debts (i.e. value of assets is less than amount of liabilities), or if it is unable to pay its debts as they fall due”, and such a company has the option to go into either (i) administration, (ii) company voluntary arrangement (CVA), (iii) receivership, (iv) liquidation or (v) dissolution. We select the failed firms from the DataStream “DeadUK” category and cross-checked at Bloomberg bankruptcy segment, Wall Street Journal (European segment) and the UK Bankruptcy & Insolvency Website for companies with at least four years firm-specific data.

2.3.2.1 Dependent variable

The dependent variable in a duration or event study is the time taken for a subject to experience an event. In our study, the event may be either a firm fails, exits a market through mergers and acquisitions, or survived beyond the sample period of the study. Therefore, the dependent variable is always stated with an event indicator, which assume value 1 for the occurrence of the event of interest and 0 otherwise. We construct the dependent variable by using the counting process input method of Andersen and Gill (1982) for the following reasons. First, it enables us to easily incorporate time varying covariates in our study. Second, the number of firms in the risk set, firms still at risk, at each period keeps changing due to constant firm entry and exit on the market. When these changes are not taken into account, less accurate estimates of the hazard rates for all our specifications may be produced. Therefore, we allow for staggered or late entry of firms and adjust the parameters for this effect. For instance, assume that it takes 5 years for a firm to be hit by an event. If the firm fails, we create the intervals $(0,1]$, $(1,2]$, $(2,3]$, $(3,4]$, and $(4,5]$ for year 1, 2, 3, 4, and 5, respectively. The event indicator is 0 for the years 1, 2, 3, and 4, when the firm is still active, but takes 1 for the 5th year, when the firm failed. We can therefore simply reconstruct the intervals as triplets: $(0,1;0]$, $(1,2;0]$, $(2,3;0]$, $(3,4;0]$, and $(4,5;1]$, where the first and second values are the beginning and the end of the year, and the last number is the event status. For example, the interval $(1,2;0]$ indicates the value of the dependent variable for the end of the second year, where 1 and 2 are the beginning and end of the second year; the third value 0 is the event indicator since the firm is still traded at the end of the second year.[4] When the firm is censored, for example, through merger and acquisition activities, we now construct the following intervals: $(0,1;0]$, $(1,2;0]$, $(2,3;0]$, $(3,4;0]$, and $(4,5;0]$. The event indicator is 0 for all the intervals since the firm left the sample as a result of a non-failure event.

2.3.2.2 Independent variables

We categorise our data set into macro-financial, firm-specific and sector level distressed indicator variables of firms. First, we employ the following macro-financial and market covariates:

(a) LSE market-wide one year trailing return, calculated by cumulating monthly market returns (Shumway, 2001). This covariate is a measure of the overall market performance, which can be used as an indicator for future performance of the UK economy.

(b) The 3 month T-bill rate, which is used as a measure of short term rates (Das et al., 2007, Duffie et al., 2007, Qi et al., 2014).

[4] In the counting process input style setting, the end of an immediate previous year is assumed to be the beginning of the next year. For the triplet $(1,2;0]$, the value 1 is the end of the first year, which is assumed to be the beginning of the second year.

Second, for the firm-specific covariates, we use two types of covariates, namely market driven and books (including balance sheets and income statements).
For market driven variables, we use:

(c) One year trailing cumulated monthly returns. The returns of stocks of distressed firms closed to a potential default are normally sold at discounted prices, making it a good hazard rate predictor (Shumway, 2001).

(d) The standard deviation of the monthly firm's equity returns: indicates how stock returns deviate from their expected value.

(e) Excess returns, computed as the difference between the stock return and the market return (Shumway, 2001, Bharath and Shumway, 2008).

(f) Market value of equity (simply equity): the product of the number of outstanding shares and current equity price (e.g. Shumway, 2001; Bharath and Shumway, 2008).

(g) Firm age: defined as the period between the time a firm is listed and the time of an event (Shumway, 2001).

(h) The face value of firm's debt, as the sum of debt in current liabilities and half of long term debt (see Vassalou and Xing, 2004.)

(i) Distance to default probability (a probabilistic measure of volatility adjusted leverage in the framework of structural model of Merton, 1974). We adopt the approach of Bharath and Shumway (2008) to construct this measure because: (i) it is much easier to implement in practice, since it does not require solving complex equations iteratively in the classical Merton's (1974) method; and (ii) it has slightly better in and out-sample predictive power, as compared to Merton's Distance-to-Default metric (Bharath and Shumway, 2008).

For book-based covariates, we employ:

(j) The ratio of net income to total assets: a measure of firm profitability over the years of operation (Zmijewski, 1984, cited in Shumway, 2001).

(k) Total liabilities to total assets, a leverage ratio, measures the firm's ability to meet it future financial obligations (Zmijewski,1984, cited in Shumway, 2001).

Third, in order to test the robustness of our model to different levels of sector distress (degrees of departure from normal market conditions), we construct five sector level distress indicators following Gilson et al. (1990), Opler and Titman (1994) and Acharya et al. (2007).[5] Let $r(t)$ be the median equity return of a sector during a given year t and $\varepsilon(n)$ be a real number that only takes on the values -0.10, -0.15, -0.20, -0.25, and -0.30 for the integer $n = 1, ..., 5$, respectively. We define a sector level distress indicator as:

$$Z(n) = \begin{cases} 1 & \text{if } r(t) < \varepsilon(n) \\ 0 & \text{otherwise.} \end{cases} \tag{2.19}$$

For example, the first sector level distress indicator is Z(1), which takes value 1 if the median equity return of a sector during a given year in the sample period of our analysis is less than -10 percent and 0 otherwise. Explicitly, this sector level distress is said to occur if the returns of over half of the number of stocks within a given sector is less than -10 percent in a particular year. The third sector level distress indicator, Z(3), corresponds to Chava et al.'s (2011) sector level distress indicator. This indicator takes value 1 if the median equity return of a sector during a given year is less than -20 percent and 0 otherwise. By our construction, the sector distress indicator 3 represents a more severe market conditions than sector distress indicator 1. All of these indicators are used to control regime changes in the sample period of our analysis.

To ensure that our results are not affected by outliers, we winsorized all the variables at 1 and 99 percentiles except distance to default probability (see e.g. Shumway, 2001; Bharath and Shumway, 2008). By construction, the distance to default probability is [0,1] bounded.

Table 2.2 presents the summary statistics of the covariates used to estimate the coefficients of the additive lognormal and multiplicative gamma frailty models in Section 2.4. The distance to default probability and sector level distress indicator have 0 and 1 as their minimum and maximum values, respectively. The stock return, market return, and excess return are bounded below and up by -91.520% and 220.557%, -22.167% and 57.840%, and -93.948% and 169.800%, respectively. The excess return has a higher standard deviation than those of the stock and market return variables. The range of the ratios net income to total assets, and the total liabilities to total assets are respectively 3.830% and 1.438%, while the former deviates more from its mean value as compared to the latter. The values of ln(age), ln(equity), ln(face value of debt) and ln(total assets) fall with the following intervals (0.000, 3.296), (11.920, 21.964), (2.606, 16.764), and (7.498, 18.368), respectively. The 3 month T-bill is bounded by (0.434%,14.332%), with not less than half of the values are more than 5.150. The lowest value of firm volatility is 0.077% while

[5] Chava et al. (2011) also followed the same authors when constructing their sector distress indicator. Here, we take a step further and construct four extra sector distress indicators.

Table 2.1: Sector names

Sector ID	Name
1	UK-DS Oil and Gas Producers
2	UK-DS Oil Equipment and Services
3	UK-DS Alternative Energy
4	UK-DS Chemicals
5	UK-DS Basic Resource
6	UK-DS Construction and Materials
7	UK-DS Aerospace and Defence
8	UK-DS General Industrials
9	UK-DS Electronic and Electrical Equipment
10	UK-DS Industrial Engineering
11	UK-DS Industrial Transportation
12	UK-DS Support Services
13	UK-DS Automobiles and Parts
14	UK-DS Food and Beverage
15	UK-DS Personal and Household Goods
16	UK-DS Health Care Equipment and Services
17	UK-DS Pharmaceuticals and Biotechnology
18	UK-DS Retail
19	UK-DS Media
20	UK-DS Travel and Leisure
21	UK-DS Fixed Line Telecommunications
22	UK-DS Mobile Telecommunications
23	UK-DS Electricity
24	UK-DS Gas, Water and Multiutilities
25	UK-DS Insurance
26	UK-DS Real Estate
27	UK-DS Financial Services(3)
28	UK-DS Software and Computer Services
29	UK-DS Technology Hardware and Equipment

Notes: We choose the 29 sub-sector due to data availability and the similarity between some of the sub-sectors. Financial Services (3) is a subsector of firms that provide financial services. This group excludes banks, real estate and insurance firms.

Table 2.2: Descriptive statistics

Variable	Mean	Std. Dev.	Min	25th P.	Median	75th P.	Max
Distance to default prob.	0.692	0.263	0.000	0.666	0.778	0.852	1.000
Stock return(%)	8.760	27.998	-91.520	-6.818	7.406	20.986	220.557
Market return (LSE)(%)	10.922	16.940	-22.167	2.590	13.170	24.080	57.840
3 month T-bill rate (%)	5.746	3.253	0.434	4.480	5.150	6.850	14.332
ln(age)	1.971	0.898	0.000	1.386	2.079	2.708	3.296
ln(equity)	16.716	1.965	11.920	15.509	16.706	17.936	21.964
Inverse of volatility	3.704	1.847	0.975	2.406	3.359	4.600	10.331
Excess return (%)	1.355	32.428	-93.948	-15.188	0.000	14.005	169.800
Stock volatility (%)	0.340	0.181	0.077	0.216	0.296	0.413	1.025
ln(face value of debt)	10.027	2.779	2.606	8.338	10.164	11.936	16.764
ln(total assets)	12.488	2.132	7.498	11.150	12.420	13.781	18.368
Total liab. to total assets	0.487	0.285	0.006	0.264	0.498	0.664	1.444
Net income to total assets	0.849	0.830	-0.055	0.090	0.683	1.298	3.825
Sector distress indicator	0.073	0.259	0.000	0.000	0.000	0.000	1.000

Notes: All covariates are winsorized at 1 and 99 percentiles, except distance to default probability covariate. The terms 25th P. and 75th P. are the 25th and 75th percentiles, respectively.

the highest value is 1.025%, whereas those of the inverse volatility are 0.975 and 10.331 respectively.

2.4 Empirical analysis

In this section, we present our empirical analysis based on the additive lognormal frailty and the multiplicative gamma frailty models using the three sets of covariates. First, we run regressions with and without firm age for the standard shared frailty and additive lognormal frailty models (see Section 2.4.1) using covariates from Duffie et al. (2007) as to examine the impact of age on firm performance in both models. Then, we employ the second set from Shumway (2001) (see Section 2.4.2). Lastly, we repeat the estimation exercise using covariates from Bharath and Shumway (2008) as to study the impact of accounting-based and market driven variables on our model and the shared frailty model. In Section 2.4.3, we compare the performance of the best model specification based on market driven covariates from Duffie et al. (2007) with that of the multiplicative gamma frailty model under various levels of market distress. We evaluate the impact of the departure from market normality on the within-sector frailties, associations and the predictive characteristics of the covariates. In Section 2.4.4, we investigate one step-ahead forecasts for the within-sector failure rates and the corresponding dependencies of our model and the multiplicative gamma frailty model by using a naïve recursive extraction approach.

Table 2.3: Additive lognormal frailty model. Dependent variable: Time to event

	Lognormal shared frailty		Additive lognormal frailty	
	M1	M2	M3	M4
Frailty variance	0.306^{**} (0.150)	0.246^{*} (0.131)	0.307^{**} (0.126)	0.288^{**} (0.147)
Additive factor			2.472^{***} (0.251)	2.422^{***} (0.250)
Distance to default prob.	1.771^{***} (0.473)	1.971^{***} (0.469)	1.703^{***} (0.467)	1.885^{***} (0.464)
Stock return	-0.017^{***} (0.003)	-0.016^{***} (0.003)	-0.015^{***} (0.003	-0.015^{***} (0.003)
Market return(LSE)	0.785^{***} (0.065)	0.786^{***} (0.065)	0.762^{***} (0.062)	0.763^{***} (0.062)
3 month T-bill rate	-1.419^{***} (0.194)	-1.373^{***} (0.194)	-0.938^{***} (0.174)	-0.897^{***} (0.174)
ln(age)		-0.392^{***} (0.103)		-0.360^{***} (0.105)
Marginal log likelihood	-632.873	-626.172	-589.175	-583.697
Likelihood ratio test	522.766*** [0.000]	531.960 *** [0.000]	610.471*** [0.000]	619.239*** [0.000]
Wald test	325.698*** [0.000]	330.184*** [0.000]	376.084*** [0.000]	382.385*** [0.000]
Pseudo-deviance	1265.746	1252.344	1178.350	1167.394
AIC	1275.746	1264.344	1190.350	1181.394
AICC	1276.070	1264.801	1190.807	1182.006
BIC	1277.151	1266.030	1192.036	1183.361

Notes: The parameter estimation is done using covariates from Duffie et al. (2007). The exact approximation is used to control for ties in the survival times of firms in our sample when deriving the penalised partial likelihood. The standard errors and p-values are in round and square brackets, respectively. The parameters are adjusted for the within-sector dependencies or correlations. AIC, AICC and BIC denote the Akaike information criterion, the corrected Akaike information criterion, and the Bayesian information criterion, respectively. *, **, and *** denote significance at 10%, 5% and 1% level, respectively.

2.4.1 Parameters estimation results using covariates of Duffie et al. (2007)

In our first specification of the model we use ln(age), distance to default probability, one year trailing market return, one year trailing stock return, and 3 month T-bill rate. The results are presented in Table 2.3. Models 1 and 2 (M1 and M2) represent the classical shared frailty models, whilst models 3 and 4 (M3 and M4) denote the additive lognormal frailty models.

The findings show that the estimated coefficient of distance to default probability is positive and statistically significant. In addition, one year trailing stock return, 3 month T-bill rate, and ln(age) are all negative and statistically significant, while one year trailing LSE stock return is unexpectedly positive and statistically significant. The frailty variance of each model (M1-M4) is a measure of the within-sector dependence or correlation between lifetimes of firms in the sectors. We argue that older firms with high stock returns are more likely to survive than younger firms with low stock returns (see e.g. Shumway, 2001). In addition, firms closer to default tends to exhibit higher probabilities of distance to default. As for the 3 month T-bill rate, the results show that this covariate tend to decrease the hazard rate. All in all, our results related to overall market are in line with those in Duffie et al. (2009) who argued that the unexpected positive sign of a market index should "not be an evidence that a good year in the stock market may in itself be bad news for default risk" (Duffie et al. 2009, page 2102). This could be attributed to the fact that in the subsequent years of a boom, a firm's distance to default probability is likely to overstate its financial prospects.

When comparing the overall fit of models M1-M4 by using AIC, $AICC$ and BIC measures, it emerges that: (i) M4 is the best model, while M1 is the worst model; (ii) the additive lognormal frailty specifications fit the data better than the share frailty models; (iii) estimates from our model may be used as early warning systems for firms.

2.4.2 Parameters estimation results using covariates of Shumway (2001) and Bharath and Shumway (2008)

The second set of covariates is taken from Shumway (2001). They are the logarithm of total assets (ln(total assets)), excess return, total liabilities to total assets, stock volatility, and net income to total assets. The results using these covariates are presented in Table 2.4. Model 5 (M5) is the classical frailty model, whilst model 7 (M7) is the additive lognormal frailty model. The estimates of these models show that the coefficients of excess return, net income to total assets and ln(total assets) are negative and statistically significant, whilst total liabilities to total assets and stock volatility are positive and statistically significant.

As for the last specification of the model, we use distance to default probability, logarithm of face value of debt (ln(face value of debt)), logarithm of equity (ln(equity)), excess return, inverse of firm volatility, and net income to total assets (see Bharath and Shumway, 2008).

In Table 2.4, model 6 (M6) is the classical shared frailty model, whereas model 8 (M8) is the additive lognormal frailty model. The results show that ln(equity), excess return, inverse of firm volatility and net income to total assets have a negative and significant impact on hazard rates, whereas ln(face value of debt) and distance to default probability covariates have a positive and significant effect on hazard rates. Again, the estimates

Table 2.4: Additive lognormal frailty model. Dependent variable: Time to event

	Lognormal shared frailty		Additive lognormal frailty	
	M5	M6	M7	M8
Frailty variance	0.200^{**} (0.100)	0.178^{**} (0.098)	0.240^{**} (0.120)	0.223^{*} (0.118)
Additive factor			2.396 (0.202)	2.393 (0.198)
Distance to default prob.		0.846^{***} (0.349)		0.715^{***} (0.358)
ln(equity)		-0.499^{***} 0.054		-0.487^{***} (0.056)
Inverse of volatility		-0.283^{***} (0.066)		-0.271^{***} (0.066)
Excess return	-0.012^{***} (0.002)	-0.009^{***} (0.002)	-0.010^{***} (0.002)	-0.006^{***} (0.002)
Stock volatility	1.748^{***} (0.330)		1.756^{***} (0.336)	
ln(face value of debt)		0.125^{***} (0.042)		0.146^{***} (0.044)
ln(total assets)	-0.318^{***} (0.040)		-0.267^{***} (0.040)	
Total liab. to total assets	0.910^{***} (0.326)		0.822^{***} (0.247)	
Net income to total assets	-0.429^{***} (0.113)	-0.305^{***} (0.111)	-0.388^{***} (0.115)	-0.274^{*} (0.113)
Marginal log likelihood	-727.920	-710.766	-659.434	-639.839
Likelihood ratio test	327.435^{***} [0.000]	360.117^{***} [0.000]	467.586^{***} [0.000]	505.431^{***} [0.000]
Wald test	353.269^{***} [0.000]	348.813^{***} [0.000]	484.286^{***} [0.000]	469.825^{***} [0.000]
Pseudo-deviance	1455.840	1421.532	1318.868	1279.678
AIC	1467.840	1435.532	1332.868	1295.678
AICC	1468.297	1436.144	1333.480	1296.469
BIC	1469.526	1437.499	1334.835	1297.926

Notes: See notes in Table 2.3.

of all the models in these specifications are adjusted for the within-sector dependencies. The two specifications, though having slightly different covariates, produce similar results.

After accounting for unobserved sector-based effects, it emerges that: (i) more profitable firms with lower debts are less likely to fail than those with less profitability and high debt; (ii) firms with higher market value are less susceptible to failure as compared to those with lower market value; and (iii) firms with high returns and relatively less volatile are likely to have higher survival rates; and (iv) firms could, to some extent, benefit from their size, since an increase in the latter decreases the instantaneous rate of failure.

In Table 2.4, we also report results for AIC, $AICC$ and BIC. The additive frailty models (M7 and M8) seems to fit the data better than the shared frailty models (M5 and M6), which collaborates the results in Section 2.4.1. These result may be informative for decision making process on the LSE.

2.4.3 Impact of sector distress on within-sector dependence

In sections 2.4.1 and 2.4.2 the information criteria measures confirm that our additive frailty model performs better than the shared frailty model. We then explore the performance of the additive lognormal frailty (ALFM) and multiplicative gamma frailty (MGFM) models under various levels of distressed market conditions in terms of data fit. We employ the five different levels of severity conditions, namely $Z(1)$, $Z(2)$, $Z(3)$, $Z(4)$, and $Z(5)$ (see Section 2.3.2). These conditions are in order of severity. For instance, the distressed market period $Z(1)$ is less severe than the distressed market period $Z(3)$.

We use the same set of covariates in specification M4 (see Table 2.3) to estimate the parameters of the additive lognormal frailty model and the multiplicative gamma frailty model, respectively, while accounting for each of the five different distressed market conditions.[6] For example, ALFM1 and MGFM1 are the additive lognormal frailty and the multiplicative gamma frailty models under distressed market condition $Z(1)$, respectively (see Table 2.5). In particular, we combine each market distress indicator with the set of covariates in specification M4 in Table 2.3, and use the new set of covariates to estimate the parameters of the additive lognormal frailty and the multiplicative gamma frailty models. For instance, for Z(1), we combine it with distance to default probability, stock return, market return, 3 month T-bill rate, and ln(age) as to form a new set of covariates that is used to estimate the additive lognormal frailty and the multiplicative gamma frailty models (see ALFM1 and MGFM1 in (Table 2.5).

The estimation results show that the coefficients of the covariates in all the regressions are similar as expected. However, the scale factor increases as the degree of severity of sector distress rises. For example, in a less severe distressed period, the estimated value of the scale factor is 1.864 for the additive lognormal frailty model (ALFM1), while this value is 2.422 in a more severe distressed period (see ALFM3). As for multiplicative gamma frailty model, the estimated values of the scale factor are 1.853 and 2.408 (see MGFM1 and MGFM3), respectively. Therefore, the scale factor for the extra variations in the hazard rates for both models increases as the market conditions becomes more severe. However, our model seems to be robust to different market conditions, as it appropriately accounts for the extra randomness induced by the distressed periods, and it performs better than the multiplicative gamma frailty model (MGFM) in measuring the within-sector dependence (see frailty variances in Table 2.5) during distressed market

[6] Model 4 is the best model amongst all the specifications and more market driven.

Table 2.5: The impact of levels of sector distress on within-sector dependence

Panel A: Additive lognormal frailty model

	ALFM1	ALFM2	ALFM3	ALFM4	ALFM5
Frailty variance	0.281^{**}	0.321^{**}	0.288^{**}	0.284^{**}	0.300^{**}
	(0.144)	(0.157)	(0.147)	(0.147)	(0.153)
Scale factor	1.864^{***}	2.159^{***}	2.422^{***}	2.851^{***}	3.218^{***}
	(0.253)	(0.250)	(0.250)	(0.256)	(0.282)
Distance to default prob.	1.924^{***}	1.895^{***}	1.885^{***}	1.839^{***}	1.809^{***}
	(0.467)	(0.465)	(0.464)	(0.463)	(0.458)
Stock return	-0.015^{***}	-0.015^{***}	-0.015^{***}	-0.014^{***}	-0.013^{***}
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
Market return	0.752^{***}	0.764^{***}	0.763^{***}	0.745^{***}	0.752^{***}
	(0.062)	(0.062)	(0.062)	(0.062)	(0.062)
3 month T-bill rate	-1.097^{***}	-0.969^{***}	-0.897^{***}	-0.823^{***}	-0.757^{***}
	(0.189)	(0.179)	(0.175)	(0.169)	(0.166)
ln(Age)	-0.364^{***}	-0.361^{***}	-0.360^{***}	-0.357^{***}	-0.354^{***}
	(0.105)	(0.105)	(0.105)	(0.106)	(0.106)
Marginal log likelihood	-598.894	-591.425	-583.697	-570.804	-563.708
Likelihood ratio test	588.594^{***}	606.143^{***}	619.239^{***}	644.525^{***}	659.174^{***}
	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]
Wald test	353.609^{***}	368.564^{***}	382.385^{***}	410.846^{***}	404.346^{***}
	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]
Pseudo-deviance	1197.788	1182.850	1167.394	1141.608	1127.416
AIC	1211.788	1196.850	1181.394	1155.608	1141.416
AICC	1212.400	1197.462	1182.006	1156.220	1142.028
BIC	1213.755	1198.817	1183.361	1157.575	1143.383

Panel B: Multiplicative gamma frailty model

	MGFM1	MGFM2	MGFM3	MGFM4	MGFM5
Frailty variance	0.220^{**}	0.257^{**}	0.228^{**}	0.225^{**}	0.220^{**}
	(0.108)	(0.122)	(0.112)	(0.111)	(0.110)
Scale factor	1.853^{***}	2.144^{***}	2.408^{***}	2.840^{***}	3.218^{***}
	(0.251)	(0.250)	(0.249)	(0.256)	(0.284)
Distance to default prob.	1.935^{***}	1.908^{***}	1.897^{***}	1.855^{***}	1.832^{***}
	(0.468)	(0.466)	(0.465)	(0.464)	(0.459)
Stock return	-0.015^{***}	-0.015^{***}	-0.015^{***}	-0.014^{***}	-0.013^{***}
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
Market return	0.751^{***}	0.763^{***}	0.761^{***}	0.743^{***}	0.752^{***}
	(0.062)	(0.062)	(0.062)	(0.062)	(0.062)
3 month T-bill rate	-1.100^{***}	-0.973^{***}	-0.900^{***}	-0.820^{***}	-0.749^{***}
	(0.190)	(0.180)	(0.175)	(0.169)	(0.166)
ln(Age)	-0.368^{***}	-0.365^{***}	-0.362^{***}	-0.358^{***}	-0.354^{***}
	(0.104)	(0.105)	(0.105)	(0.106)	(0.106)
Marginal log. likelihood	-797.804	-784.520	-776.700	-763.991	-756.789
Likelihood ratio test	586.363^{***}	604.063^{***}	616.886^{***}	642.677^{***}	656.858^{***}
	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]
Wald test	350.689^{***}	365.668^{***}	379.738^{***}	409.226^{***}	400.407^{***}
	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]
Pseudo-deviance	1595.608	1569.040	1553.400	1527.982	1513.578
AIC	1609.608	1583.040	1567.400	1541.982	1527.578
AICC	1610.220	1583.652	1568.012	1542.594	1528.190
BIC	1611.575	1585.007	1569.367	1543.949	1529.545

Notes: See notes in Table 2.3.

periods.

For robustness of analysis, we also estimate the within-sector failure rates (frailties) and random effects (log-frailties)(see section 2.1) using our model, ALFM3, and model of Chava et al. (2011), MGFM3. The results are presented in Figure 2.1 (see panels A and B).

It emerges that firms in sectors with frailties greater than one tend to fail faster than firms with frailties less than one. For instance, firms in the Real Estate sector (see sector ID. 26 in Table 2.1) with a frailty of 1.918 for ALFM3 (1.676 for MGFM3) are likely to fail faster than firms in the Fixed Line Telecommunications sector (see Sector ID. 21 in Table 2.1) with a frailty of 0.907 for ALFM3 (0.890 for MGFM3). Therefore, these figures confirm the results in Table 2.5, and they seem to suggest that under distressed market periods, the additive lognormal frailty model is likely to outperform the multiplicative gamma frailty model.

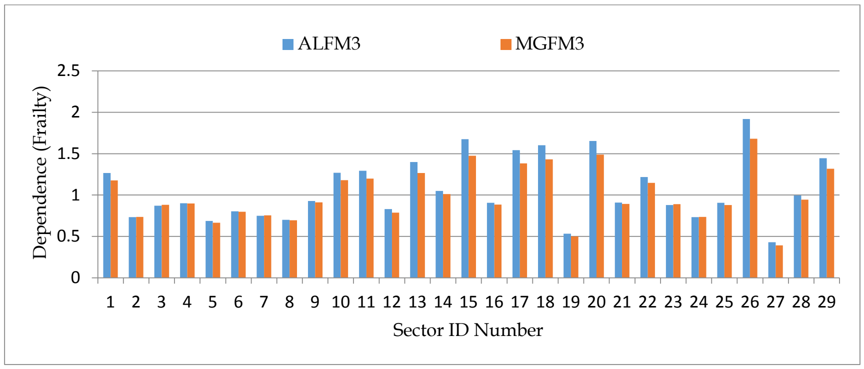
As for the overall model fit of ALMF (ALFM1-ALFM5) and MGFM (MGFM1-MGFM5) specifications, we compare ALFM1 to MGFM1, ALFM2 to MGFM2, and so on, using AIC, $AICC$ and BIC. It is shown that ALFM1 is preferred to MGFM1, when the median stock return is less than -10% in distress periods, ALFM2 outperforms its counterpart MGFM2, when the median stock return is less than -15%. This implies that the additive frailty model may offer more accurate information on changes in hazard rate driven by various levels of distress severity on the LSE than the multiplicative frailty model.

2.4.4 Out-of-sample extraction of failure rates

The accuracy of the estimates of failure rates plays a central role in stakeholders' decisions. In this section we use an out-of-sample parameter extraction approach to extract sector-level failure rates (frailties are not observable). We present the results of one step-ahead extracts by using our model and the multiplicative gamma frailty model. More, specifically we consider one-year horizon, as often required by most regulatory requirements (see for instance the Bank for International Settlements), and compute the additional deviations from the expected future values. We then evaluate the accuracy of the extraction by using the root mean square of the deviations: the higher the value of this metric, the higher the accuracy.

We proceed as follows. We use a naive recursive scheme for one-step ahead extraction over the following years: 2010, 2011 and 2012. For instance, to extract the within-sector frailty (or sector-level failure rate) and the corresponding dependence for 2010, we define a sample from 1985 to 2010 and estimate the parameters using the period 1985

Panel A: Within-Sector Dependences



Panel B: Within-Sector Random Effects

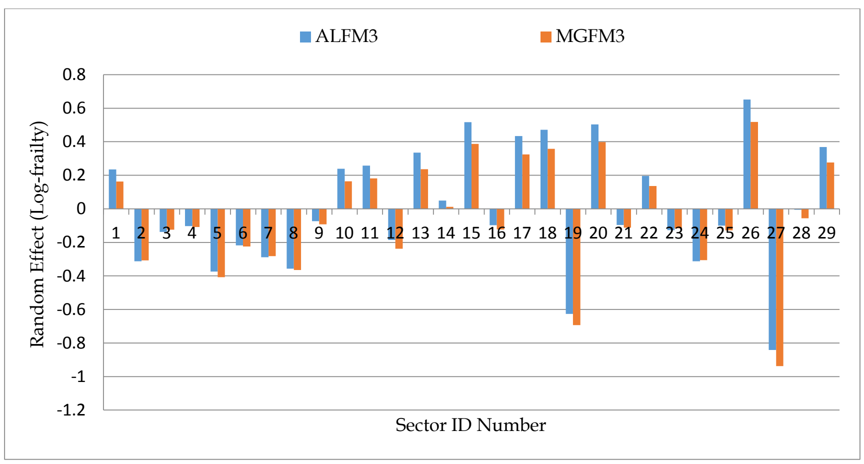


Figure 2.1: Estimated failure rates and log-failure rates for models ALFM3 and MGFM3.

- 2009 by holding out 2010. In this way, we obtain the frailties at the beginning of 2010. We do the same for 2011 and 2012. This naive extraction scheme is repeated for all the sectors under consideration. Finally, for each sector i, we construct the root mean square deviation ($RMSD_i$) as follows:

$$RMSD_i = \sqrt{\sum_{t=1}^{3} (\hat{y}_{i,t} - E(\tilde{u}_{i,t}))^2/3}, \tag{2.20}$$

where $\hat{y}_{i,t}$ denotes the extracted value, $E(\tilde{u}_{i,t})$ is the expected value of frailty for sector $i = 1, ..., 29$ and $t = 1, 2, 3$, where $t = 1$ indicates the year 2010, $t = 2$ is the year 2011, $t = 3$ is the year 2012.[7] We use the expected value of the frailty as the actual value since it is not observable at the end of 2010. For instance, $\hat{y}_{5,2}$, the extracted failure rate for the UK Basic Resource sector (ID.5) for the year 2011. Table 2.6 presents the additive lognormal model and multiplicative gamma frailty model extractions based on the within-sector dependencies (see Table 2.7). Results for the RMSD are illustrated in Table 2.8.

The results in Table 2.6 show that there are differences in the extracted values over time and across sectors for both models. This extraction allows us to distinguish between firms in sector which are likely to fail faster or slower in the event of firm failure clustering. Firms in sectors with estimates larger than 1 (fast-failure regime) are likely to fail faster, whilst those with estimates smaller than 1 (slow-failure regime) are likely to fail slower. For instance, firms in the UK Oil and Gas Production Sector (ID. 1) are likely to fail faster, while those in the UK Health Equipment and Services sector (ID. 16) are likely to fail slower. Furthermore, these results reveal some interesting trends in firm failure. First, in a fast-failure regime, the multiplicative gamma frailty model tends to underestimate these rates across sectors, while the additive lognormal frailty tends to predict these rates more accurately. For instance, for the UK Real Estate Sector (ID. 26), the extractions of the failure rates for the multiplicative model are 1.469, 1.631 and 1.676, whereas those for the additive lognormal frailty model are 1.753, 1.853 and 1.918, respectively. In addition, these dynamics also hold for a mixed regime, where firms are likely to fail slower in some years and faster in others (see e.g. sector ID. 14). Second, in the slow-failure periods, the multiplicative gamma frailty tends to overestimate the rates, whilst additive lognormal frailty model predicts (extracts) these rates more accurately. For example, for the UK Alternative Energy Sector (ID. 8), the predictions of the multiplicative gamma model for the rates are 0.979, 0.906 and 0.882 and those of the additive model are 0.966, 0.893 and

[7] The impact of frailties on hazard rates during distressed periods tends to be more pronounced and hence we construct a metric for capturing the additional variations in hazard rates across the years for each sector. Therefore, high values of our metric are desirable.

Table 2.6: Within-sector failure rate extractions, $\hat{y}_{i,t}$.

	ALFM			MGFM		
Sec. ID	2010	2011	2012	2010	2011	2012
1	1.437	1.246	1.313	1.235	1.163	1.210
2	0.848	0.746	0.727	0.894	0.754	0.730
3	0.966	0.893	0.870	0.979	0.906	0.882
4	0.974	0.904	0.899	0.976	0.901	0.893
5	0.733	0.682	0.656	0.772	0.663	0.630
6	0.832	0.801	0.801	0.869	0.799	0.795
7	0.778	0.749	0.744	0.837	0.757	0.748
8	0.788	0.732	0.699	0.845	0.733	0.693
9	0.889	0.924	0.933	0.906	0.909	0.913
10	1.015	1.297	1.285	0.990	1.201	1.188
11	1.052	1.244	1.299	1.020	1.163	1.200
12	0.640	0.797	0.800	0.653	0.760	0.757
13	1.147	1.417	1.424	1.079	1.274	1.279
14	0.964	1.014	1.044	0.961	0.983	1.003
15	1.471	1.655	1.680	1.285	1.459	1.473
16	0.739	0.952	0.924	0.800	0.930	0.899
17	1.364	1.533	1.540	1.210	1.377	1.379
18	1.375	1.604	1.667	1.231	1.432	1.474
19	0.813	0.534	0.519	0.830	0.503	0.483
20	1.595	1.626	1.687	1.397	1.467	1.510
21	1.147	0.927	0.906	1.078	0.914	0.890
22	1.300	1.243	1.224	1.159	1.164	1.149
23	0.910	0.883	0.878	0.941	0.895	0.888
24	0.824	0.759	0.731	0.875	0.768	0.735
25	0.949	0.928	0.915	0.951	0.903	0.886
26	1.735	1.853	1.918	1.469	1.631	1.676
27	0.451	0.432	0.425	0.462	0.394	0.384
28	1.121	1.011	0.992	1.058	0.963	0.940
29	1.411	1.333	1.441	1.236	1.232	1.311

Notes: The reported estimates denotes the failure rate with an expected value of 1. In the event of failure clustering, firms with estimates larger (lower) than 1 are likely to failure faster (slower). These estimates are adjusted for the within-sector dependencies or correlations in Table 2.7.

Table 2.7: Out-of-sample within-sector dependence extracts

	ALFM			MGFM		
Sec. ID	2010	2011	2012	2010	2011	2012
1	0.234	0.202	0.196	0.123	0.152	0.148
2	0.339	0.334	0.358	0.206	0.334	0.369
3	0.392	0.401	0.428	0.206	0.332	0.367
4	0.307	0.313	0.335	0.177	0.265	0.288
5	0.218	0.193	0.205	0.166	0.205	0.224
6	0.268	0.285	0.307	0.178	0.268	0.291
7	0.312	0.336	0.366	0.207	0.334	0.369
8	0.315	0.297	0.308	0.207	0.305	0.331
9	0.240	0.259	0.278	0.159	0.222	0.238
10	0.197	0.181	0.191	0.130	0.137	0.146
11	0.266	0.266	0.287	0.155	0.188	0.199
12	0.132	0.112	0.118	0.115	0.108	0.114
13	0.361	0.371	0.393	0.176	0.217	0.233
14	0.251	0.235	0.250	0.157	0.193	0.203
15	0.144	0.140	0.149	0.089	0.101	0.107
16	0.297	0.228	0.228	0.207	0.195	0.202
17	0.186	0.144	0.150	0.109	0.106	0.112
18	0.131	0.118	0.121	0.086	0.089	0.091
19	0.176	0.155	0.151	0.133	0.182	0.177
20	0.088	0.082	0.082	0.060	0.065	0.065
21	0.286	0.261	0.274	0.155	0.223	0.240
22	0.291	0.271	0.285	0.146	0.191	0.205
23	0.366	0.396	0.432	0.206	0.332	0.367
24	0.329	0.340	0.361	0.206	0.334	0.370
25	0.246	0.201	0.210	0.156	0.177	0.188
26	0.103	0.087	0.089	0.066	0.066	0.068
27	0.138	0.099	0.102	0.140	0.115	0.117
28	0.150	0.127	0.132	0.102	0.112	0.118
29	0.187	0.140	0.136	0.108	0.110	0.106

Notes: The estimates represent the dependence or correlation between the lifetimes of firms in the sectors.

0.870, respectively. The results of our model seem to offer a more accurate classification of firms in terms of failure speed, and they may be useful for an appropriate portfolio reshuffling.

Table 2.8: Root mean square deviations

Sec ID	ALFM3	MGFM3
1	0.313	0.210
2	0.273	0.270
3	0.130	0.118
4	0.101	0.107
5	0.344	0.370
6	0.199	0.205
7	0.256	0.252
8	0.301	0.307
9	0.067	0.087
10	0.285	0.188
11	0.299	0.200
12	0.200	0.243
13	0.424	0.279
14	0.044	0.003
15	0.680	0.473
16	0.076	0.101
17	0.540	0.379
18	0.667	0.474
19	0.481	0.517
20	0.687	0.510
21	0.094	0.110
22	0.224	0.149
23	0.122	0.112
24	0.269	0.265
25	0.085	0.114
26	0.918	0.676
27	0.575	0.616
28	0.008	0.060
29	0.441	0.311

When comparing the RMSD of the two models for each sector, the additive lognormal frailty model averagely has slightly higher values than those by the multiplicative gamma frailty model (see Table 2.8). These findings seem to confirm the relevance of our distribution assumption on the frailties, as, on average, the additive lognormal frailty model fits the data better than the multiplicative gamma frailty model during distressed market periods.

2.5 Conclusions

We use a multivariate lognormal regime-switch frailty model to estimate and predict within-sector failure rates and the corresponding dependencies of listed firms on the London Stock Exchange (LSE) over period 1985-2012. The model is particularly suitable for dealing with distressed market periods. In relation to a set of observable predictive factors of failure rates, we find significant evidence of unobserved sector-specific source of default rates amongst the listed firms. Neglecting these unobserved sector-specific factors may likely lead to underestimation of the hazard rates.

We also account for an adjustment factor in hazard rates and investigate the dynamics of this relative to a set of crucial firm failure predictive factors when moving away from normal market conditions. The scale adjustment increases when moving from less to more severe distressed market conditions, whilst the desirable impact of distance to default probability (volatility adjusted leverage) with a substantial predictive power for hazard rates averagely deteriorates. However, all the other covariates also experience slight changes in their magnitudes as expected. Interestingly, we also found that the distance to default probability of firms is likely to overstate the financial prospects of these firms after a boom on the LSE.

We also compare our model with the multiplicative gamma frailty model of Chava et al. (2011). It results that the former outperforms the latter both in-sample and out-sample estimates, as it offers much flexibility in accounting for extra variations in hazard rates induced by departure from market normality and unobserved sector factors. The outcomes in terms of goodness-of-fit are confirmed when using information criteria. We argue that the additive lognormal frailty is likely to produce better estimates and predictions of hazard rates, within-sector failure rates and dependencies.

Our findings have some important implications for stakeholders on the LSE. Specifically, in the event of failure clustering on the LSE, the within-sector failure rates of our model could be used by investors and other stakeholders to discriminate amongst firms or sectors, which are likely to fail faster or slower. In this respect, investors may effectively rebalance their portfolios and obtain good estimates of their portfolio risks. On the other hand, regulators may rank firms into various risk profiles in order to suitably design new or enhance existing regulatory requirements to make firms more risk sensitive. Further, since the hazard rate specification heavily depends on distance to default probability covariate, market participants are highly recommended not to be conservative on firms' distance to default probability after a market boom on the LSE. Failing to account for this may likely lead to underestimation of default rates, within-sector failure rates and dependencies of firms.

Appendix A

A.1 Definition and derivation

Let $T \in [0, \infty)$ be the event time, which is a continuous random variable with the probability density function $f(t)$ and a corresponding cumulative distribution function $F(t)$ as follows:

$$F(t) = P(T \leq t) = \int_0^t f(t)dt \tag{A.1}$$

The survival function $S(t)$ beyond time t is given by:

$$\begin{aligned} S(t) = P(T > t) &= 1 - F \\ &= 1 - \int_0^t f(t)dt, \end{aligned} \tag{A.2}$$

where $S(t)$ is a continuous and strictly decreasing function since T is a continuous random variable. Furthermore, we define hazard rate or function (also called failure rate) as follows:[1]

$$h(t) = \lim_{\Delta \to \infty} \frac{P(t < T \leq t + \Delta | T > t)}{\Delta t} \tag{A.3}$$

Given that T is a continuous random variable, equation (A.3) can be rewritten as

$$h(t) = \frac{f(t)}{S(t)} = -\frac{dS(t)}{dt} \tag{A.4}$$

[1] For detailed treatment, refer to Duchateau and Janssen, 2008; Wienke, 2011; Hangal, 2011; Kleinbaum and Klein, 2012; and others.

We derive an associated measure called cumulative hazard function by integrating both sides of equation (A.4) as follows:

$$H(t) = \int_0^t h(t)dt = -lnS(t). \tag{A.5}$$

Again, the survival function can be derived by integrating both sides of equation (A.5)

$$S(t) = exp(-H(t)) = exp(-\int_0^t h(x)dx) \tag{A.6}$$

A.2 Cox proportional hazard model

Let T_i and δ_i be the event time and event indicator, respectively, for firm i. The term δ_i assumes value 1 when T_i is a failure time and 0 otherwise. The Cox proportional hazard model for firm i is given by:

$$h_i(t) = h_0(t)exp(X_i(t)\beta), \tag{A.7}$$

where $h_0(t)$ is an unspecified term called the baseline hazard rate, $X_i(t)$ is the set of covariate for firm i with p-dimensional parameter estimates. Equation (A.7) can either be a parametric, semi-parametric or non-parametric model depending on how the the baseline hazard rate $h_0(t)$ is treated. We obtain a parametric model if we impose a distribution (see e.g. exponential, Weibull, and gamma distributions) on the baseline hazard rate, while in the case of semi-parametric the latter is left unspecified. In equation (A.7), for the non-parametric models, no distribution is assumed, and hence one need to apply numerical techniques which normally require larger samples. In this thesis, we formulate all our models in the semi-parametric context. Specifically, we employ the Cox semi-parameter model and its extended form (see e.g Therneau and Grambsch, 2000 for a detailed and excellent treatment) in estimating the parameters in all the specifications in the thesis.

Cox (1972, 1975) proposed the partial likelihood approach, that do not depend on the baseline hazard rate as follows:

$$L_i(\beta) = \prod_{i=1}^{n} \left(\frac{exp(X(T_i)\beta)}{\sum_{j \in R(T_i)} exp(X(T_i)\beta)} \right)^{\delta_i}, \tag{A.8}$$

where $R(T_i)$ is the risk set (individuals at risk at time T_i). Taking natural logarithm of equation (A.8), we obtain:

$$l_i(\beta) = \sum_{i=1}^{n} \delta_i \left(X(T_i)\beta - \log \left(\sum_{j \in R(T_i)} exp(X(T_i)\beta) \right) \right), \tag{A.9}$$

A.3 Classical shared frailty model

Consider a sample of s clusters and in each cluster there are n_i members. In addition, let $(T_{ij}, \delta_{ij}, X_{ij}(t))$ be the observed clustered data, where $(T_{ij},\ \delta_{ij})$, $X_{ij}(t)$ are the event time, censoring indicator and a set of covariates respectively for member j in cluster i. Following McGilchrist and Aisbett (1991), we restate equation (A.7) in the context of shared frailty as

$$h_{ij}(t) = h_0(t)exp(X_{ij}(t)\beta + w_i), \tag{A.10}$$

where $w_i = logu_i$ is a random effect or log-frailty term shared by all members in cluster i. Equation (A.10) can be restated as

$$h_{ij}(t) = h_0(t)u_i exp(X_{ij}(t)\beta), \tag{A.11}$$

where $u_i = exp(w_i)$.

Further, we present the full and partial likelihoods as follows.

A.3.1 Full maximum likelihood

Following the classical approach of maximum likelihood (ML) in the literature, we construct the conditional ML for cluster i as follows:

$$L_i(\beta; X_{ij}(t)) = \prod_{j=1}^{n} h(X_{ij}(t), T_{ij}; \beta)^{\delta_{ij}} S(X_{ij}(t), T_{ij}; \beta), \tag{A.12}$$

where $h(.)$ and $S(.)$ are hazard and survival functions, respectively.

From equation (A.11), we can rewrite equation (A.12) as:

$$L_i(\psi, \beta; X_{ij}(t)) = \prod_{j=1}^{n} h_0(t)u_i exp(X_{ij}(t)\beta)^{\delta_{ij}} exp(-H_0(t)u_i exp(X_{ij}(t)\beta)), \tag{A.13}$$

where ψ is the parameter(s) of the baseline hazard, and $H(T_{ij})$ is the cumulative hazard function. To solve the likelihood in equation (A.13), we impose an appropriate distribution on the frailty term u_i. The gamma and lognormal distribution are widely used in the literature. The gamma distribution has closed form expressions for the likelihood, while the lognormal do not and hence one would have to use numerical approximations. However, the lognormal distribution offers much more flexibility as compared to gamma distribution in the multivariate frailty context (Hougaard, 2000; Klein and Moeschberger, 2003; Duchateau and Janssen, 2008; Wienke, 2011; Hanagal, 2011; and others). To estimate the full likelihood, sum over all the s clusters.

A.3.2 Penalised partial likelihood

The random effect is assumed to follow a normal distribution on a log-scale. We write the likelihood which is conditioned on the random effect term w_i for cluster i as:

$$\begin{aligned} L_i(\beta,\theta;w_i) &= \prod_{j=1}^{n_i} f(X(T_{ij},\delta_{ij},w_i;\beta,\theta) \\ &= \prod_{j=1}^{n_i} f(X(T_{ij},\delta_{ij},w_i;\beta) * f(w_i;\theta) \\ &= l_{i1}(\beta;w_i) * l_{i2}(\theta;w_i). \end{aligned} \tag{A.14}$$

The firm term, $l_{i1}(\beta;w_i)$ is the partial likelihood, $l_{i2}(\theta;w_i)$ is the penalty term, and θ is the variance of the random effect terms (see Duchateau and Janssen, 2008).

The terms $l_{i1}(\beta;w_i)$ and $l_{i2}(\theta;w_i)$ are respectively given by:

$$L_{i1}(\beta) = \prod_{i=1}^{n_i} \left(\frac{exp(X(T_{ij})\beta) + w_i}{\sum_{j\in R(T_{ij})} exp(X(T_{ij})\beta) + w_j} \right)^{\delta_{ij}}, \tag{A.15}$$

and

$$L_{i2}(\theta;w_i) = f_i(\theta;w_i) = \frac{1}{\sqrt{2\pi\theta}} exp(-\frac{w_i^2}{2\theta}). \tag{A.16}$$

Equation (A.15) is obtained by restating equation (A.8) in the context of clustered data, and accounting for random effect. Furthermore, taking log of equation (A.14) we obtain the following:

$$l_i(\beta,\theta;w_i) = logL_i(\beta,\theta;w_i) = logl_{i1}(\beta;w_i) + logl_{i2}(\theta;w_i)$$
$$= \sum_{i=1}^{n} \delta_i \left(X_i(t)\beta + w_i - \log\left(\sum_{j\in R(T_i)} exp(X_i(t)\beta + w_j) \right) \right) - \frac{1}{2}\left(\frac{w_i^2}{\theta} + log(2\pi\theta) \right). \tag{A.17}$$

There are s number of clusters, and to obtain the penalised partial likelihood for all cluster, we sum equation (A.17) over s clusters:

$$l(\beta,\theta;w) = \sum_{i=1}^{s}\sum_{i=1}^{n} \delta_i \left(X_i(t)\beta + w_i - \log\left(\sum_{j\in R(T_i)} exp(X_i(t)\beta + w_j) \right) \right) - \frac{1}{2}\sum_{i=1}^{s} \left(\frac{w_i^2}{\theta} + log(2\pi\theta) \right), \tag{A.18}$$

Therefore

$$l_{ppl}(\beta,\theta;w_i) = l_{part}(\beta;w) - l_{pen}(\theta;w), \tag{A.19}$$

where

$$l_{part}(\beta;w) = \sum_{i=1}^{s}\sum_{i=1}^{n} \delta_i \left(X_i(t)\beta + w_i - \log\left(\sum_{j\in R(T_i)} exp(X_i(t)\beta + w_j) \right) \right), \tag{A.20}$$

and

$$l_{pen}(\theta;w) = \frac{1}{2}\sum_{i=1}^{s} \left(\frac{w_i^2}{\theta} + log(2\pi\theta) \right). \tag{A.21}$$

Equations (A.19), (A.20) and (A.21) are the penalised partial likelihood, partial part and penalised term, respectively.

Chapter 3

Bias correction in hazards rates: Evidence from USA default corporate data

3.1 Introduction

A firm exit event can be relevant to investment making decision and regulatory exercise. Firms may leave the market because of mergers and acquisitions (see Draper and Paudyal, 2006; Baker et al., 2012), failure to rebalance leverage (Pour and Lasfer, 2013), weak corporate governance structures (Marosi and Massoud, 2008; Hostak et al., 2013, and orders). In such cases, market participants can grasp relevant information on the general performance of firms, and gauge the trade-off between risk and returns of a portfolio. Therefore, it is extremely important to account for firm censoring when estimating default rates of firms, since neglecting informative censoring may produce bias in the estimation of the hazard rates.

This study proposes a novel approach that accounts for informative firm censoring and unobserved factors in order to estimate the hazard rates of public listed firms on the NYSE, NASDAQ, and NYSE MKT LLC (AMEX). In this respect, the approach here differs from previous ones (see Dewaelheyns and Van Hulle, 2006; Duffie et al., 2007; Bharath and Shumway, 2008; Duffie et al., 2009; Chava et al, 2011; Jacobson et al., 2013; Qi et al., 2014; Atsu and Costantini, 2015; among others) which assume that non-default firm exit provides no explicitly relevant information, and has no impact on the hazard rate of active firms. We draw on the assumption that high or low risk of default firms are normally censored, with the frequency of censoring higher and its impact more pronounced during distressed periods. Thus, we employ an inverse censoring probability weighted scheme (see Robins, 1993; Robins and Finkelstein, 2000; Scharfstein and Robins, 2002).

We proceed as follows. First, we compute the censoring probability by using firm specific factors, macroeconomic conditions, and industry-level unobserved factors for each period a firm spends in the active group across all the industries. Second, we derive the survival probabilities from the censoring probabilities, and construct the weights as the inverse of the survival probabilities. In this fashion, the firms, which are likely to be censored, are given higher weights than their counterparts, since such firms normally have lower survival probabilities. Finally, we specify the maximum likelihood function of hazard rate using the weights in order to correct for the bias in the estimates.

This study contributes to literature in two respects. First, we construct weights to deal with potential effects of censoring, and these weights vary with changes in firm specific and macroeconomic factors. We assume that reasons for censoring are the same across all the industries. Further, we explicitly account for these weights in estimating the hazard rates and industry level failure rates of the surviving firms. In this respect, we are able to capture potential shocks in the covariates which hardly can be handled by a classical dummy variable approach. As a result, our approach adjusts the default rates up and is more likely to produce accurate portfolio risk estimates as compared to the non-informative censoring approaches. Second, since censoring activities may vary across industries, due to diverse industry characteristics (see e.g. Andrade et al., 2001; Harford, 2005), we construct weights that combine firm specific and macroeconomic factors with industry level activities. We derive our models by using these weights to correct the potential bias in the estimates of the multiplicative frailty models of Chava et al. (2011) and the additive frailty model of Atsu and Costantini (2015), which are used as benchmark models.

We compare in and out-of sample performances of our bias-corrected models with those of Chava et al. (2011) and Atsu and Costantini (2015) in order to give insight into the estimation of default rates during distress market periods. We use distance to default probability (see Bharath and Shumway, 2008), one year trailing S&P 500 return, one year trailing stock return (see Shumway, 2001), 3 month T-bill rate (see Duffie et al., 2007) and an industry distress indicator which is used to account for the extra variations in hazard rates during distressed market periods.

Our main findings are as follows. First, an increase in one year trailing stock return and 3 month T-bill rate cause the hazard rates to decrease. While, hazard rates tend to increase with an increase in distance to default probability and one year trailing S&P 500 return in all the models. An increase in hazard rates decreases the expected time to default of firms and vice versa. Second, the significant and positive sign of the distressed indicator in both classical and our bias-corrected models provides the evidence of high hazard rates during distressed periods, and hence these rates should be adjusted accordingly. Our models perform the adjustment accurately than the benchmark models, since the latter underestimate the scale of adjustment. Third, the significance of the industry level frailty factor supports the importance of multivariate frailty models that adjusts hazard rates for industry-level factors. Our technique accurately measures the industry level correlation induced by these factors than the benchmarks. Lastly, the standard errors in our bias-corrected models are lower than those of the benchmark models.

We also examine the accuracy of a one step-ahead forecasts of industry level failure rates (or frailties) and dependencies of our bias-corrected and the benchmark models using a naive recursive extraction approach (see Chapter 2 for the details). We evaluate the performance of these models using the root mean square deviations. The industry level failure rates with expected value of 1 are used to discriminate among those industries which firms are likely to fail faster or slower in the event of failure clustering. More specifically, firms which show failure rates lower than 1 tend to fail slowly, while those with failure rates more than 1, are likely to fail faster. We perform the extraction exercise over the time horizon 2009-2013. The results show that: (i) during slow failing periods, the benchmark models tend to produce higher values of failure rates than the bias-corrected models; (ii) in a faster failing periods, the benchmarks generate lower values of failure rates as compared to the bias-corrected models.

In general, our bias-corrected models seem to perform better than the benchmark models in-sample and out-of-sample estimation exercises.

The rest of the study is organized as follows. Section 3.2 presents the methodology and data. Specifically, in Section 3.2.1, we briefly present the benchmark models, and use these models to motivate and formulate our bias-corrected models in Section 3.2.2. In addition, we present our data set and pre-estimation results for the various industries in our sample. Section 3.3 presents the empirical results for all models, and Section 3.4 concludes the study.[1]

3.2 Methodology and data

In this section, we first briefly present the benchmark models, namely the multiplicative frailty (MF) model of Chava et al. (2011) and the additive frailty (AF) model of Atsu and

[1] This chapter heavily draws on the literature of Chapter 2, and hence we do not review literature again.

Costantini (2015).[2] These models assume that censoring of firms on the market conveys no relevant information to investors. This may produce potential bias in the estimates of hazard rates, especially during distressed market periods. Then, we present our models, which are derived from correcting the potential bias in the benchmark models by using an inverse probability of censoring weighted (IPCW) scheme (see Robins, 1993; Robins and Finkelstein, 2000; Scharfstein and Robins, 2002). We consider the classical and the adjusted schemes. The classical scheme assumes that reasons for censoring are the same across the industries, while in the adjusted scheme censoring turn to be diverse across industries. Lastly, we describe the data set used for the empirical analysis.

3.2.1 Benchmark models

In each model, there are s industries and n_i firms in each industry. We consider the following sample period $[0, T^*]$. Then, firms may: (i) enter in the sample in a staggered manner; (ii) leave the sample through non-failure or failure events; and (iii) survive beyond the end of the period, T^*. The information set is $(T_{ij}, \delta_{ij}, X_{ij}(t), \tilde{u}_i(t))$, for $i = 1, ..., s$ and $j = 1, ..., n_i$, where $X_{ij}(t)$ is the vector of time-varying covariates for firm j in industry i, T_{ij} denotes the event time, δ_{ij} is the censoring indicator, and $\tilde{u}_i(t)$ indicates an industry frailty term. The term δ_{ij} takes value 1 if T_{ij} is failure time and 0 otherwise. The model of Chava et al. (2011) is given by:

$$h_{ij}(t) = \begin{cases} u_i \Delta^{Z_i(t)} exp(X_{ij}(t)\beta) & \text{if sector } i \text{ is distressed,} \\ u_i exp(X_{ij}(t)\beta) & \text{otherwise,} \end{cases} \tag{3.1}$$

where β represents a p-dimensional vector of regression parameters, u_i is the frailty term for a specific industry, $Z_i(t)$ is a time-varying industry level distress indicator, which assumes 1 for distressed industries (with a multiplicative factor Δ) at time t and 0 otherwise. The combination of u_i and $Z_i(t)$ gives the term $\tilde{u}_i(t)$. In order to estimate the parameters in equation (3.1), the frailty term for all the industries is assumed to follow a gamma distribution given by:

$$f(u) = \frac{u^{1\theta} - 1}{\theta^{1/\theta}\Gamma(1/\theta)} exp(-u/\theta) \tag{3.2}$$

where $A_i(t) = 1/\theta(t) + \sum_{j=1,T_{ij}<t}^{n_{ij}} \delta_{ij}$ and $C_i(t) = 1/\theta(t) + \sum_{j=1,T_{ij}<t}^{n_{ij}} H(T_{ij})$ indicate the shape and scale parameters, respectively, and $H(T_{ij}) = \int_0^{T_{ij}} (\Delta^{Z_i(t)} exp(X_{ij}(t)\beta) dt$. In the

[2] For details of these models, refer to Chapter 2.

above setting, Chava et al. (2011) constructed the marginal likelihood for all industries as follows:

$$l(\theta, \Delta, \beta) = \sum_{i=1}^{s} l_i(\theta, \Delta, \beta), \tag{3.3}$$

where the likelihood for each industry is given by:

$$\begin{aligned} l_i(\theta, \Delta, \beta) = \log \Gamma(\delta_i + 1/\theta) - \log \Gamma(1/\theta) - (1/\theta)\log(\theta) + \sum_{j=1}^{n_i} \delta_{ij}(X_{ij}(T_{ij})\beta \\ + Z_i(T_{ij})\log(\Delta)) - (\delta_i + 1/\theta)\log(1/\theta + \sum_{j=1}^{n_i} H(T_{ij})), \end{aligned} \tag{3.4}$$

and the gamma function, $\Gamma(.)$, has a mean of 1 and a finite variance θ.

The model of Atsu and Costantini (2015) is given by the following:

$$h_{ij}(t) = \begin{cases} exp(X_{ij}(t)\beta + \pi Z_i(t) + w_i) & \text{if sector } i \text{ is distressed,} \\ exp(X_{ij}(t)\beta + w_i) & \text{otherwise,} \end{cases} \tag{3.5}$$

where w_i is the log-frailty with mean 0 and variance γ, $\pi = log(\Delta)$ denotes the additive factor and $Z_i(t)$ is the industry distress indicator. Atsu and Costantini (2015) derived a penalised partial likelihood in the spirit of McGilchrist and Aisbett (1991) as follows:

$$l_p(\beta^*, \gamma | w) = \sum_{i=1}^{s} \sum_{j=1}^{n_i} \delta_{ij} \left(X\beta^* + w_i - \log \left(\sum_{j \in R(T_{ij})} exp(X\beta^* + w_j) \right) \right) - \frac{1}{2\gamma} \sum_{i=1}^{s} w_i^2, \tag{3.6}$$

where $\beta^* = (\beta, \pi)$ is a $(p+1)$-dimensional vector of regression parameters, and $R(T_{ij})$ is the set of all firms still at risk of a default. In order to make the parameter estimation feasible, Atsu and Costantini (2015) applied the extension and approximation of Ripatti and Palmgren (2000) to derive equation (3.6) and equation (3.7) as below:

$$l_m(\beta^*, \gamma) = -\frac{1}{2}\log(\gamma I) + \log(\int exp[l_p(\beta^*, \gamma)]dw), \tag{3.7}$$

and

$$l_m(\beta^*, \gamma) \approx -\frac{1}{2}\log(\gamma I) + \log(|H_{22}(\beta^*, \gamma, w^*)|) - l_p(\beta^*, \gamma, w^*), \tag{3.8}$$

where I is an identity matrix of order $s \times s$, where s is the number of industries in the sample, and H is the negative Hessian.

3.2.2 Our approach

In our setting, firm censoring is regarded as an absorbing state, i.e. once a firm has left the sample through either a failure or non-failure event, it does not return to it. Further to this assumption, we present our approach based on the use of two weighting schemes within the context of the extended Cox proportional hazard model. In the first scheme ($k = 1$), termed as classical scheme, we adjust the censoring probabilities for only firm-specific and macroeconomic covariates. While in the second scheme ($k = 2$), which we call industry level adjusted scheme, we account for the industry-based unobserved factors in addition to the firm specific and macroeconomic factors.

Let $C_{ij}^k(t)$ denote the censoring probability of firm j in industry i for k type of scheme at time t, which is defined as follows:

$$C_{ij}^k(t) = \begin{cases} C_{0ij}^k(t)exp(X_{ij}(t)\beta) & \text{for } k = 1, \\ C_{0ij}^k(t)\tilde{u}_i(t)exp(X_{ij}(t)\beta) & \text{for } k = 2, \end{cases} \tag{3.9}$$

where $C_{0ij}^k(t)$ is the baseline hazard function. Specifically, the censoring probability is conditioned on the time-varying covariates, $X_{ij}(t)$ for $k = 1$, while conditional on time-varying covariates and the unobserved factor, $\tilde{u}_i(t)$, for $k = 2$. For instance, $C_{ij}^2(t)$ is the censoring probability conditional on firm specific, macroeconomic, and unobserved industry level factors.

Further, we deduce a survival function for the two types as follows:

$$S_{ij}^k(t) = \begin{cases} [S_{0ij}^k(t)]^{exp(X_{ij}(t)\phi)} & \text{for } k = 1, \\ [S_{0ij}^k(t)]^{exp(X_{ij}(t)\phi+U)} & \text{for } k = 2, \end{cases} \tag{3.10}$$

where $S_{ij}^k(t)$ is the survival probability for firm j in industry i at time t, and $S_{0ij}^k(t) = exp(-\int_0^T C_{0ij}^k(t))dt$ is the baseline survival probability for type k. The term U is set to zero for the classical scheme, while it is equal to $\tilde{u}_i(t)$ for the industry level adjusted scheme. Finally, we construct the weights by using the survival probability:

$$W_{ij}^k(t) = \frac{1}{S_{ij}^k(t)}, \tag{3.11}$$

where $0 < S_{ij}^k(t) < 1$. We can deduce from equation (3.11) that greater weights are allotted to firms who have lower survival probability beyond time t for any of the 2 schemes. In other words, such firms exhibit higher probability of censoring. We perform bias correction in equations (3.3) and (3.6) using the estimated dynamic weights in equation (3.11). In this respect, the contribution of each quarterly data point of the firms to the maximum likelihoods is weighted accordingly. Equation (3.3) can be then written as :

$$l(\theta^*, \Delta^*, \beta_w) = \sum_{i=1}^{s} l_i(\theta^*, \Delta^*, \beta_w), \tag{3.12}$$

where each industry likelihood is given by

$$\begin{aligned} l_i(\theta^*, \Delta^*, \beta_w) = \log\Gamma(\delta_i + 1/\theta^*) - \log\Gamma(1/\theta^*) - (1/\theta^*)\log(\theta^*) + \sum_{j=1}^{n_i} \delta_{ij}(X_{ij}(T_{ij})\beta_w \\ + Z_i(T_{ij})\log(\Delta^*)) - (\delta_i + 1/\theta^*)\log(1/\theta^* + \sum_{j=1}^{n_i} H^*(T_{ij})), \end{aligned} \tag{3.13}$$

where, β_w, Δ^* and θ^* are the bias-corrected coefficients, multiplicative factor and frailty variance respectively.[3]

Likewise, equation (3.6) becomes:

$$l_p(\beta_w^*, \gamma^*|w) = \sum_{i=1}^{s}\sum_{j=1}^{n_i} \delta_{ij}\left(X\beta_w^* + w_i - \log\left(\sum_{j\in R(T_{ij})} exp(X\beta_w^* + w_j)\right)\right) - \frac{1}{2\gamma^*}\sum_{i=1}^{s} w_i^2, \tag{3.14}$$

where equations (3.7) and (3.8) respectively become:

$$l_m(\beta_w^*, \gamma^*) = -\frac{1}{2}\log(\gamma^* I) + \log(\int exp[l_p(\beta_w^*, \gamma^*)]dw), \tag{3.15}$$

and

$$l_m(\beta_w^*, \gamma^*) \approx -\frac{1}{2}\log(\gamma^* I) + \log(|H_{22}(\beta_w^*, \gamma^*, w^*)|) - l_p(\beta^*, \gamma, w^*), \tag{3.16}$$

where β_w^* , and γ^* are the bias-corrected estimates. In Atsu and Costantini's (2015) setting, the parameters estimation is done in log-frailties, and we therefore derive the frailty variance, θ^*, from log-frailty variance, γ^*, using the relationship: $\gamma^* = log(\theta^* + 1)$ (see Duchateau and Janssen, 2008). We term the bias-corrected models of MF and AF as weighted multiplicative frailty model (WMF) and weighted additive frailty model (WAF), respectively. Further, CWAF and CWMF are the bias-corrected models for the models

[3] For brevity, the bias corrected shape and scalar parameters, $A_i^*(t)$ and $C_i^*(t)$ are not presented here.

AF and MF using the classical weights, while the models AWAF and AWMF are obtained using the adjusted weights. The bias-corrected estimates, β^*, β^*_w, Δ^*, θ^*, and γ^*, are expected to be more efficient as compared to their non-weighted counterparts, β, β^*, Δ, θ, and γ. This stems from the fact that the bias correction scheme may use a richer information set, which may include informative firm censoring, and explicitly adjusts these estimates for potential sharp changes in the covariates of firms.

When all the weights, $W^k_{ij}(t)$, are equal to 1, the WMF and WAF models collapse into models MF and AF, respectively.

3.2.3 Data

Our sample data is drawn from five main sources, namely Center for Research in Security Prices (CRSP), COMPUSTAT, Moody's Default and Recovery Database (DRD), the UCLA-LoPucki Bankruptcy Research Database (BRD), and Board of Governors of the Federal System over the period 1980-2013 for the NYSE, NASDAQ and NYSE MKT LLC (AMEX) exchanges.

We employ the Global Industry Classification Standard (GICS) since it is widely used by financial practitioners and also classifies firms consistently over the years as compared to other classification systems (see Bhojraj et al., 2003).[4] The GICS is a "four-tiered, hierarchical industry classification system" (MSCI, 2015 page 2), which comprises of 10 sectors (first hierarchy), 24 industry groups (second hierarchy), 67 industries (third hierarchy) and 156 sub-industries (fourth hierarchy). In order to examine the within-industry frailties and dependence structures of firms in the same line of business activities, we use the fourth hierarchy and exclude the financial and utility firms. We consider 127 industries in total.

We merge market driven variables from monthly CRSP file and accounting variables from quarterly COMPUSTAT file using PERMNO number. A PERMNO number is a primary key that uniquely identifies a stock within the CRSP setting. For the analysis, we use event time as the dependent variable. It represents the time until a firm experiences either a failure or non-failure event. We use the Andersen and Gill (1982) counting process style of input in dealing with time-varying variables and staggered firm entry (for details, see Chapter 2). The independent variables are: Distance to default probability (DDP), one year trailing stock return of a firm (see Shumway, 2001), one year trailing S&P 500 return

[4]These classification systems broadly include (1) Standard Industrial Classification (SIC); (2) North American Industry Classification System (NAICS); and (3) Fama and French (1997) algorithm. Readers can refer to Bhojraj et al. (2003) for a detailed comparison of these systems.

(see Shumway, 2001), 3 month T-bill rate, and an industry level distress indicator.[5] DDP is a probabilistic measure of volatility adjusted leverage, and firms with higher (smaller) probability are closer (farer) to default. In constructing this metric, we follow Bharath and Shumway (2008). The 3 month T-bill rate is a measure of risk free rate. Finally, to construct the industry distress indicator, we follow Chava et al. (2011). The indicator takes value 1 if the median of the stock returns of firms is less than -20 percent in a given industry during a quarter, and 0 otherwise.

For the definition of default, we follow the literature (see Duffie et al., 2007; Das et al, 2007; Qi et al., 2014; Azizpour et al., 2015, among others), and adopt the Moody's definition, which includes the following cases: missed interest payment, distressed exchange offers, reorganization, and bankruptcy. On the basis of this definition, we have 1350 failed firms and 7121 active firms in our data set, which is equal to 417,295 quarterly firms observations. For the active firms, 4048 are censored (left the sample before the end of the study period) due to voluntary or involuntary delisting reasons, and 3073 survived beyond the study period as shown in Table 3.1 below.

Table 3.1: Sector level statistics of active and failed firms

Sector ID	Sector name	Active firms	Failed firms	Total
10	Energy	593	115	708
15	Materials	517	93	610
20	Industrials	1189	230	1419
25	Consumer Discretionary	1352	294	1646
30	Consumer Staples	345	64	409
35	Health Care	1182	210	1392
45	Information Technology	1754	302	2056
50	Telecommunication Services	189	42	231
	Grand total	7121	1350	8471

Notes: For brevity sake, we present firm status distribution at the GICS sector level. However, we use 127 industries in our empirical analysis for all the models (see Table B.1 in Appendix B).

In Table 3.2, we present the descriptive statistics for all the covariates employed in our empirical analysis. The DDP is bounded by [0.000, 1.000], with about three quarters of the firms' probabilities less than 0.410. Further, the mean of the stock return and the S&P500 return are approximately the same (i.e 0.700 percent), but the stock return deviates more from its expected value than the market index. The 3 month T-bill rate ranges between

[5] To construct the DDP we use as inputs: market value of an equity, as the product of share price at the end of a quarter and the number of outstanding shares, stock return, stock volatility (monthly realised volatilities that are scaled with the number of trading days in a given month, see Shumway, 2001), the face value of debt, as the sum of debt in current liabilities and half of long term debt, see Vassalou and Xing, 2004.

Table 3.2: Descriptive statistics

Variable	Mean	Std Dev.	Min.	25th P.	Median	75th P.	Max.
DDP	0.230	0.186	0.000	0.034	0.221	0.410	1.000
Stock return	0.706	12.220	-98.440	-4.750	0.033	5.423	158.140
S&P 500 return	0.702	4.850	-13.202	-2.138	1.198	3.895	13.391
3 month T-bill rate	4.483	3.233	0.010	1.750	4.830	5.900	16.310
Distressed indicator	0.007	0.082	0.000	0.000	0.000	0.000	1.000

Notes: All covariates are winsorized at 1 and 99 percentiles, except DDP. Stock return, S&P 500 return and 3 month T-bill rate are expressed in percentages. The terms 25th P. and 75th P. are the 25th and 75th percentiles, respectively.

0.001 and 16.310, with an expected value and standard deviation of 4.483 and 3.233, respectively. As expected, the industry level distress indicator jumps to 1 for distressed periods, while it stays or reverts to 0 during normal market conditions. Overall, the stock return and S&P 500 return tend to exhibit higher variations than the other variables.

3.3 Empirical analysis

In Section 3.2.2, we argue that the WAF and WMF models are likely to produce more accurate estimates of failure rates as compared to the benchmark models, AF and MF. In this section, we empirically explore the forging argument by comparing the performance of our model with the benchmarks using the US data. We combine a set of covariates from Shumway (2001), Duffie et al. (2007), and Bharath and Shumway (2008) with a regime switch variable.

3.3.1 Parameter estimates of the benchmark models

For the benchmark models, we use the following variables: DDP, one year trailing stock return, one year trailing S&P 500 return, 3 month T-bill rate, and a distress indicator at industry level.[6] The parameter estimates of the benchmark models are presented in Table 3.3. AF1-AF4 represent different specifications of the additive frailty model, while MF1-MF4 indicate different multiplicative frailty models. Models AF1 and MF1 are the univariate specifications, where the hazard rate functions depend only on the distance to default probability, which turns to be highly significant. In models AF2 and MF2, we control for distressed market periods. The distance to default probability is still significant, and the distressed indicator is also significant. In the other model specifications, we combine

[6] For details, see also Chapter 2.

Table 3.3: Benchmark models. Dependent variable: Time to event

	AF				MF			
Variables	AF1	AF2	AF3	AF4	MF1	MF2	MF3	MF4
Scale factor		0.805^{***} (0.258)		0.859^{***} (0.264)		0.803^{***} (0.259)		0.861^{***} (0.265)
DDP	2.080^{***} (0.156)	2.072^{***} (0.156)	1.592^{***} (0.158)	1.585^{***} (0.158)	2.086^{***} (0.156)	2.077^{***} (0.156)	1.604^{***} (0.158)	1.598^{***} (0.159)
Stock return			-0.034^{***} (0.002)	-0.034^{***} (0.003)			-0.034^{***} (0.002)	-0.033^{***} (0.003)
S&P 500 return			0.370^{***} (0.010)	0.371^{***} (0.010)			0.374^{***} (0.010)	0.375^{***} (0.010)
3 month T-bill rate			-0.481^{**} (0.244)	-0.481^{**} (0.243)			-0.495^{**} (0.244)	-0.495^{**} (0.246)
Diagnostics								
Industry level								
Frailty variance	0.145^{***}	0.146^{***}	0.337^{***}	0.408^{***}	0.152^{***}	0.154^{***}	0.343^{***}	0.416^{***}
Global								
ML	-10701	-10696	-10901	-10086	-12051	-12046	-11438	-11433
LR test	399.440	408.728	1688.333	1698.707	402.9341	412.2976	1972.656	1710.840
	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]
Wald test	333.383	343.612	1968.661	1972.007	333.998	344.297	1700.355	1975.813
	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]

Notes: The efron approximation is used to adjust for ties in the event times of firms. The standard errors and p-values are in round and square brackets, respectively. *,**, and *** denote significance at 10%, 5% and 1% level. MR and LR indicate the marginal log-likelihood and the Likelihood ratio, respectively. In all the models, the frailty variance is a measure of industry level dependence between the event times.

the distance to default probability and other covariates. The distance to default probability retains its significance. In addition, stock return and 3 month T-bill rate are negative and statistically significant at 1% and 5% levels, respectively, while S&P 500 return is positive and significant at 1% level. In all the models, the frailty variance is statistically significant, supporting the importance of incorporating industry level dynamics in estimating hazard rates. Further, the failure dependence for each model falls in the range (0.145, 0.416). Besides, the scale factor is positive and significant. This points to the importance of accounting for higher variations in hazards rates under a departure from market normality.

The results suggest the following. First, the hazard rate rises with an increase in the distance to default probability and S&P covariates leading to a decrease in the expected time to default. Second, an increase in the stock return and 3 month T-bill rate variables drive the hazard rate down causing an increase in the expected time to default.

3.3.2 Parameter estimates of the weighted models.

Using the same set of covariates as in Section 3.3.1, we now present the estimates of our bias-corrected models. Table 3.4 presents the bias-corrected models using the classical weights. CWAF1-CWAF4 are the bias-corrected models for additive frailty model, while models CWMF1-CWMF4 for the multiplicative frailty model. This implies that model

CWAF1 is a bias-corrected model obtained from model AF1 by controlling for informative firm censoring using the classical weights. Likewise model WMFC1 is obtained from MF1, and so on. The following results emerge. First, the distance to default probability is highly significant in both the univariate (models CWAF1 and CWMF1) and multivariate specifications (models CWAF2, CWAF3, CWAF4, CWMF2, CWMF3, and CWMF4). Second, all the covariates retain their significance levels and the expected signs, except the 3 month T-bill rate whose significance level has improved from 5% to 1%. The correlation or failure dependence for each model falls within the interval (0.287, 0.594). In Table 3.5, we present the bias-corrected models using the industry level factors adjusted weights. In a similar fashion as in Table 3.4 , we obtain model AWAF1 from model AF1 by adjusting the parameter estimates for the potential effects of informative censoring using the industry level factors adjusted weights. Likewise model AWMF1 is obtained from model MF1, and so on. The results are similar to those of the results of the bias-corrected models using classical weights, although there are marginal differences. In these models, the failure rate dependencies fall within the interval (0.240, 0.577).

Table 3.4: Bias-corrected models using classical weights. Dependent variable: Time to event

	CWAF				CWMF			
Variables	CWAF1	CWAF2	CWAF3	CWAF4	CWMF1	CWMF2	CWMF3	CWMF4
Scale factor		1.559^{***}		1.633^{***}		1.565^{***}		1.645^{***}
		(0.144)		(0.147)		(0.145)		(0.147)
DDP	2.284^{***}	2.260^{***}	1.857^{***}	1.841^{***}	2.290^{***}	2.267^{***}	1.867^{***}	1.860^{**}
	(0.130)	(0.130)	(0.131)	(0.131)	(0.130)	(0.130)	(0.131)	(0.132)
Stock return			-0.029^{***}	-0.029^{***}			-0.029^{***}	-0.029^{***}
			(0.002)	(0.002)			(0.002)	(0.002)
S&P 500 return			0.383^{***}	0.387^{***}			0.387^{***}	0.392^{***}
			(0.009)	(0.009)			(0.009)	(0.010)
3 month T-bill rate			-0.720^{***}	-0.715^{***}			-0.733^{***}	-0.735^{***}
			(0.200)	(0.199)			(0.201)	(0.200)
Diagnostics								
Industry level								
Frailty variance	0.287^{***}	0.296^{***}	0.528^{***}	0.568^{***}	0.360^{***}	0.370^{***}	0.564^{***}	0.597^{***}
Global								
ML	-18024	-17977	-17189	-16966	-19356	-19308	-18515	-18291
LR test	792.499	889.347	2516.028	2618.052	808.400	905.489	2533.113	2634.836
	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]
Wald test	683.082	802.152	2865.200	2910.435	687.267	806.229	2869.462	2911.307
	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]

Notes: See Table 3.3 for notes.

3.3.3 Comparison of parameter estimates

In this section, we first compare the results of the estimated parameters for the benchmark and bias-corrected models. Then, using the estimated parameters, we calculate and com-

Table 3.5: Bias-corrected models using adjusted weights.Dependent variable: Time to event

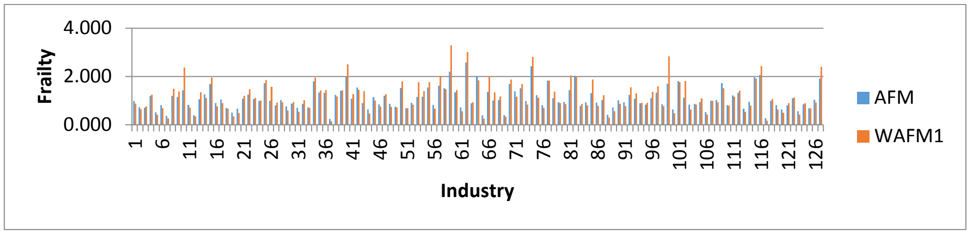
	AWAF				AWMF			
Variables	AWAF1	AWAF2	AWAF3	AWAF4	AWMF1	AWMF2	AWMF3	AWMF4
Scale factor		1.323^{***} (0.157)		1.404^{***} (0.160)		1.330^{***} (0.157)		1.410^{***} (0.161)
DDP	2.041^{***} (0.119)	2.025^{***} (0.119)	1.590^{***} (0.121)	1.578^{***} (0.121)	2.048^{***} (0.120)	2.032^{***} (0.120)	1.600^{***} (0.121)	1.588^{***} (0.121)
Stock return			-0.032^{***} (0.002)	-0.032^{***} (0.002)			-0.032^{***} (0.002)	-0.032^{***} (0.002)
S&P 500 return			0.379^{***} (0.008)	0.381^{***} (0.008)			0.383^{***} (0.008)	0.384^{***} (0.008)
3 month T-bill rate			-0.603^{***} (0.187)	-0.602^{***} (0.186)			-0.616^{***} (0.187)	-0.614^{***} (0.186)
Diagnostics								
Industry level								
Frailty variance	0.240^{***}	0.246^{***}	0.516^{***}	0.535^{***}	0.328^{***}	0.335^{***}	0.560^{***}	0.577^{***}
Global								
MR	-20314.8	-20286.1	-19249.1	-19218.3	-21641.2	-21612.3	-20567.2	-20536.1
LR test	654.653	816.934	2957.958	3023.010	778.415	838.265	2977.291	3042.222
	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]
Wald test	757.292	726.071	3486.745	3523.869	663.776	735.090	3491.251	3526.754
	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]

Notes: See Table 3.3 for notes.

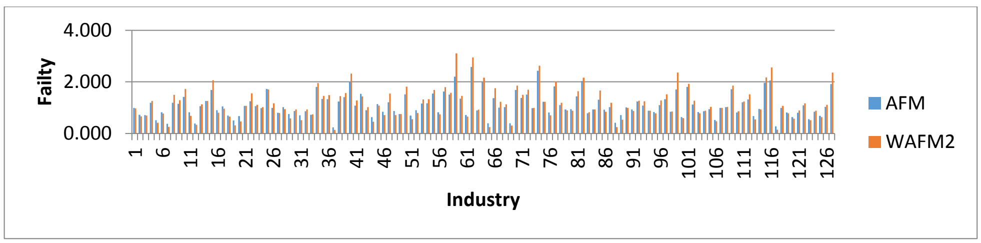
pare the industrial level failure rates among the two models. The following trends emerge. First, the 3 month T-bill rate plays a higher significant role in estimating failure rates in the bias-corrected models than in the benchmarks. Second, the impact of unobserved factors on hazard rates, which induces failure dependence, is better capture by our models as higher values of failure dependencies are found for these models. Third, our models tend to estimate the scale factor more appropriately than the benchmark models because of higher values of the factor during distressed market periods. In conclusion, while the classical models are very conservation, which may not be ideal during distressed market conditions, the bias-corrected models are more forward looking: adjust failure rates accordingly during booms and unfavourable market periods.

In Figure 3.1, results concerning the industry level failure rates estimated using both the benchmark and bias-corrected models are illustrated. In particular, firms in industries with frailty greater one (see e.g. Fertilizers & Agricultural Chemicals with ID. 10 and GICS code 15101030, Table B.1) are likely to fail faster, while those firms with frailty less than one (see e.g. Railroads with ID 44 with GICS code 20304010) will fail slower. Further, the figure shows that, during distressed market periods, usually characterized by higher failure rates and firm exits, the benchmark models relatively generate lower failure rate. In addition, during relatively normal market periods, usually featured by lower failure rates, the bias-corrected models are able to generate lower values for the failure than the benchmark models.

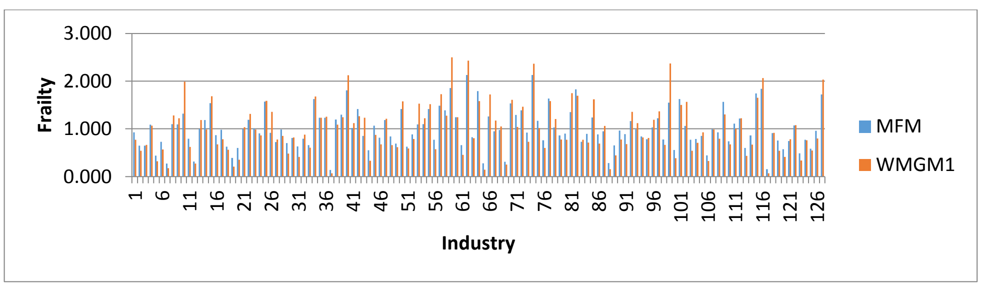
Panel A: Industry Level Frailties for AFM and WAFM1



Panel B: Industry Level Frailties for AFM and WAFM2



Panel C: Industry Level Frailties for MFM and WMFM1



Panel D: Industry Level Frailties for MFM and WMFM2

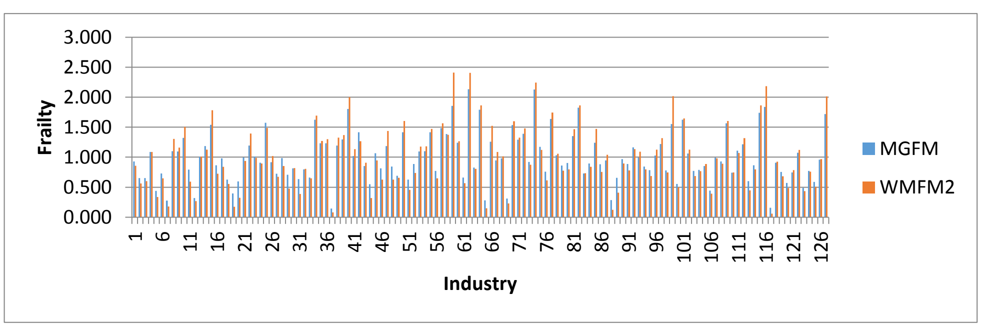


Figure 3.1: Industry level failure rates

The results in Sections 3.3.1-3.3.3 have some implications. First, the distance to default probability shows a high explanatory power for the hazard rate, and it may also be used as early warning system by financial investors who can efficiently rebalance their portfolios by disposing of firms with a high and/or consistent increase in their distance to default probabilities, while investing more in those firms with a lower distance to default probability. Second, the negative and significant effect of 3 month T-bill rate shows how this covariate is likely to have a dampen effect on failure rates on the US firms. The regulatory authorities and investors, for instance, can exploit this dampen effect to their advantage. However, during distressed market periods, which are normally characterised by more firm exits, the effects of the 3 month T-bill rate can be captured more appropriately, when firm censoring is explicitly accounted for in the failure rates. Third, the unexpected positive effect of market return on hazard rates (see also Duffie et al., 2007; Duffie et al., 2009) may offer investors a general overview of firm performance after market booms. As such, investors can use important trends during market booms on the US markets to reassess risk levels of firms in the subsequent periods. Fourth, the significance of the sector level distress indicator shows that hazard rates of firms are adjusted appropriately when those firms belong to distressed industries in a given period. In particular, this can help an investor re-calculate the risk and return of his portfolio containing firms of the distressed industries.

3.3.4 Out-of-sample extraction of industry level failure rates

Accurate extracts of failure rates and the corresponding failure rate dependence can be used to rank industries into risky and less risky ones. In this section, we employ a naive technique (see Atsu and Costantini, 2015) to extract industry level failure rate of the benchmarks, AF and MF, and the bias-corrected models, WAF and WMF, and compare the performances of the two for a one year horizon in terms of accuracy. To this end, we compute the root mean square deviations (RMSD) for each industry as deviations of the extracts from their expected future values. Higher values of this measure imply higher accuracy. Indeed, during booms or distressed market periods, industry level failure rates tend to deviate more from their expected values.[7]

We proceed with the recursive extraction exercise over the years 2009, 2010, 2011, 2012, and 2013 as follows. To extract one-step-ahead industry level failure rate for 2009, we choose the sample period 1980- 2009. We then estimate the parameters of the benchmarks and the bias-corrected models over the 1980-2008 period, and use estimates to

[7] For details, see Chapter 2.

extract rates for the beginning of 2009. This year is regarded as the out-of-sample period. The RMSD is computed as follows for each industry $i = 1, ..., 127$:

$$RMSD_i = \sqrt{\frac{\sum_{t=1}^{5}(\hat{y}_{i,t} - E(\tilde{u}_{i,t}))^2}{5}}, \tag{3.17}$$

where $\hat{y}_{i,t}$ and $E(\tilde{u}_{i,t})$ are the extracted and expected values respectively for $t = 2009, ..., 2013$, $t = 1$ represents the year 2009, $t = 2$ is the year 2010, and so on. For example, $\hat{y}_{8,1}$ denotes the extracted failure rate for the Commodity Chemicals industry (ID 8 and GICS code 15101010) for 2009. As for the extraction exercise, we use the second weighting schemes (see Section 3.2.2) and the results are presented in Tables 3.6 and 3.7.

First, in Table 3.6, some of the industries have failure rates less than 1 (see e.g. Oil & Gas Equipment & Services: GICS code 10101020) over the entire extraction horizon, 2009-2013. In this industry, the one-step-ahead predictions of the industry level failure rates in 2009 for the models AF and WAF are equal to 0.482 and 0.400, respectively. In 2010, these rates increased to 0.741 and 0.687 for the models AF and WAF, respectively, and over the period 2011-2013, the rates vary between 0.771-0.811 for AF model and 0.660-0.757 for WAF model. As for the prediction of the MF model, over the years 2009-2013, the extracted values fall within the ranges 0.365-0.723, while those of the WAF model are bounded between 0.282 and 0.629 (see Table 3.7). The variations in the industry level failure rates are driven by changes in macroeconomic, firm specific and industry level factors. Specifically, the deterioration (improvement) of these factors may push up (down) the rates. When the extracted rates are all less than one over the extraction time horizon, the industries show a low risk profile, a slow failing period is observed, and firms in these industries are likely to fail slowly in the event of failure clustering. Second, other industries have failure rates greater than 1 (see e.g. Fertilizers & Agricultural Chemical: GICS code 15101030), and show a high risk profile, with a fast failing period over the time horizon. Third, a few industries have failure rates less than 1 in some years and greater than 1 in some others (see e.g. Construction Materials: GICS code 15102010). In such a case, industries show a mixed risk profile, and with a mixed failing period.

As for the three different failing periods (slow, fast and mixed), some patterns emerge. First, during the period of slow failing, the AF and MF models tend to overestimate the industry level failure rates as compared to the AWAF and AWMF models, respectively. For example, the extracted values of the Silver industry (GICS code: 15104045) in 2009 are equal to 0.691 and 0.479 for AF and AWAF models, respectively. In other words, model AF suggests that firms in the industry are 31% less risky, whereas for model AWAF the firms are 52% less risky.[8] Similar percentages are also found for the two models AF and AWAF over the period 2010-2013. On average, the benchmark models produce estimates of the industry level failure rates larger than those of bias-corrected models by 0.200 over the period 2009-2013. Second, during the fast failing regime, the AF and MF models tend to underestimate the failure rates as compared to the AWAF and AWMF models. For instance, in 2009, firms in the Paper Packaging industry (GICS code: 15103020) have failure rate equal to 1.395 for the MF model as opposed to 1.507 for the AWMF model. This implies that firms in the Paper Packaging industry are approximately 40% and 51% more risky in 2009 when applying the two models. Overall, the benchmarks models produce estimates of the failure rates lower than those by the bias-corrected models. Lastly, during the mixed risk period, all the models exhibit both the trends in the fast and slow failing periods (see e.g. the Paper Products industry 15105020).

The results concerning the accuracy of the extraction exercise over the period 2009-2013 are reported in Table 3.9. On average, the AWAF and AWMF models produce higher RMSD values than the AF and MF models during the slow failing period. For instance, in the Oil & Gas Refining & Marketing industry (ID 5 and GICS code 10102030), AF and AWAF generate values of 0.4830 and 0.584 for the RMSD, respectively, while the RMSD is 0.624 for the MF model, and 0.736 for the AWMF. The same outcome is observed in the fast failing period (see e.g. ID 10 and GICS code 15101030) and the mixed failing period (see ID 15 and GICS code 15103020). All in all, the AWAF and AWMF averagely generate higher RMSD values than those of AF and MF for all the three periods.

3.4 Conclusion

In this study, we correct for potential bias in the multiplicative and additive frailty models of Chava et al. (2011) and Atsu and Costantini (2015) using an inverse probability censoring weighted scheme (see Robins 1993; Robins and Finkelstein, 2000; Scharfstein and Robins, 2002). In this way, the bias-corrected models allow us to explicitly account for informative firm censoring. We assume that censored firms normally belong to either good or bad performing firms categories. For instance, two performing firms may benefit from forming a synergy, or a financially distressed firm may be willing to be taken over by a healthy firm. In both cases, the target firm is censored, and this may convey useful information to the market participants, with a potential impact on hazard rates of active firms.

[8] These percentages are obtained by subtracting the failure rates from the expected value, which is equal to 1.

Table 3.6: Failure rate extracts using AF and AWAF

		AF					AWAF				
ID.	GICS code	2009	2010	2011	2012	2013	2009	2010	2011	2012	2013
1	10101010	0.977	0.910	0.975	1.043	0.986	0.943	0.865	0.931	1.042	0.967
2	10101020	0.482	0.741	0.811	0.771	0.722	0.400	0.687	0.757	0.709	0.660
3	10102010	0.949	0.850	0.797	0.764	0.710	1.117	0.943	0.836	0.776	0.693
4	10102020	1.345	1.365	1.142	1.156	1.192	1.514	1.510	1.221	1.228	1.262
5	10102030	0.448	0.461	0.510	0.442	0.517	0.313	0.341	0.386	0.327	0.416
6	10102040	0.338	0.531	0.748	0.755	0.816	0.264	0.457	0.692	0.694	0.773
7	10102050	0.214	0.293	0.410	0.369	0.371	0.123	0.192	0.290	0.250	0.255
8	15101010	0.922	0.921	1.002	1.137	1.188	1.229	1.174	1.257	1.451	1.498
9	15101020	1.548	1.387	1.252	1.204	1.149	1.877	1.655	1.469	1.390	1.290
10	15101030	2.211	1.843	1.539	1.367	1.420	2.910	2.444	1.965	1.689	1.726
11	15101040	0.769	0.796	0.825	0.828	0.823	0.621	0.651	0.681	0.684	0.682
12	15101050	0.241	0.273	0.393	0.366	0.388	0.198	0.224	0.347	0.310	0.338
13	15102010	1.224	1.106	1.017	0.980	1.059	1.393	1.219	1.086	1.025	1.133
14	15103010	1.379	1.498	1.349	1.304	1.253	1.351	1.531	1.379	1.327	1.253
15	15103020	1.653	1.464	1.500	1.426	1.687	1.972	1.722	1.809	1.701	2.063
16	15104010	1.016	0.973	0.942	0.923	0.897	1.002	0.939	0.872	0.840	0.802
17	15104020	1.334	1.316	1.104	1.095	1.049	1.242	1.250	0.992	1.005	0.957
18	15104030	0.792	0.772	0.832	0.690	0.693	0.792	0.760	0.827	0.639	0.648
19	15104040	0.601	0.634	0.666	0.503	0.506	0.411	0.440	0.457	0.301	0.312
20	15104045	0.691	0.707	0.719	0.703	0.668	0.479	0.502	0.503	0.486	0.460
21	15104050	1.421	1.354	1.172	1.169	1.072	1.508	1.440	1.190	1.185	1.069
22	15105010	1.591	1.447	1.323	1.285	1.247	2.117	1.896	1.713	1.644	1.555
23	15105020	1.275	1.133	1.032	0.988	1.063	1.399	1.205	1.065	0.999	1.105
24	20101010	1.041	0.998	0.909	1.005	0.980	1.102	1.042	0.933	1.058	1.024
25	20102010	2.533	2.215	1.940	1.842	1.726	2.587	2.240	1.962	1.851	1.705
26	20103010	1.115	1.065	1.014	1.036	0.988	1.441	1.332	1.238	1.249	1.165
27	20104010	0.808	0.876	0.943	0.803	0.799	0.759	0.862	0.967	0.790	0.790
28	20104020	1.370	1.239	1.135	1.077	1.018	1.346	1.214	1.105	1.020	0.935
29	20105010	0.772	0.763	0.771	0.769	0.756	0.602	0.592	0.596	0.592	0.581
30	20106010	0.843	0.867	0.900	0.851	0.870	0.961	0.958	0.981	0.906	0.931
31	20106015	0.626	0.658	0.701	0.705	0.697	0.450	0.479	0.517	0.519	0.517
32	20106020	1.097	0.978	0.918	0.870	0.862	1.250	1.091	1.014	0.948	0.925
33	20107010	0.769	0.687	0.652	0.789	0.718	0.897	0.751	0.667	0.849	0.744
34	20201010	2.229	1.970	1.886	1.778	1.799	2.391	2.106	2.072	1.940	1.954
35	20201050	1.444	1.484	1.420	1.353	1.332	1.597	1.637	1.577	1.489	1.455

Notes: For the sake of brevity, we present only the first 35 industries, but interested readers are referred to Table B.2 in Appendix B.2 for the full table. We present the second type bias correction estimates of frailty rates. These frailties, with an expected value of 1, are used to discriminate among industries-firms which ones are likely to fail faster or slower. Firms in industries with frailty greater than one are inclined to fail faster, while those with frailty less than 1 are likely to fail slowly in the event of failure clustering. We use the frailties as industry level failure rates.

Table 3.7: Failure rate extracts using MF and AWMF

		MF					AWMF				
ID.	GICS code	2009	2010	2011	2012	2013	2009	2010	2011	2012	2013
1	10101010	0.846	0.821	0.910	0.980	0.929	0.739	0.717	0.806	0.913	0.855
2	10101020	0.365	0.621	0.723	0.691	0.651	0.282	0.535	0.629	0.596	0.561
3	10102010	0.831	0.769	0.735	0.705	0.650	0.882	0.788	0.724	0.676	0.601
4	10102020	1.105	1.181	1.027	1.049	1.088	1.142	1.212	1.027	1.047	1.088
5	10102030	0.317	0.354	0.419	0.357	0.440	0.205	0.244	0.294	0.250	0.335
6	10102040	0.229	0.417	0.646	0.660	0.727	0.172	0.338	0.559	0.568	0.647
7	10102050	0.112	0.188	0.307	0.270	0.276	0.060	0.117	0.204	0.171	0.176
8	15101010	0.759	0.799	0.912	1.045	1.098	0.931	0.947	1.070	1.247	1.303
9	15101020	1.314	1.247	1.178	1.142	1.095	1.437	1.362	1.284	1.232	1.159
10	15101030	1.782	1.594	1.413	1.275	1.321	2.142	1.937	1.669	1.462	1.506
11	15101040	0.725	0.756	0.789	0.795	0.792	0.532	0.562	0.587	0.596	0.593
12	15101050	0.159	0.192	0.316	0.292	0.318	0.128	0.157	0.269	0.241	0.269
13	15102010	1.064	1.007	0.957	0.927	1.005	1.093	1.019	0.953	0.909	1.012
14	15103010	1.191	1.336	1.259	1.228	1.185	1.067	1.269	1.208	1.177	1.125
15	15103020	1.395	1.308	1.383	1.329	1.538	1.507	1.412	1.552	1.481	1.782
16	15104010	0.940	0.925	0.909	0.892	0.867	0.845	0.827	0.785	0.759	0.722
17	15104020	1.138	1.172	1.026	1.022	0.981	0.966	1.030	0.855	0.875	0.840
18	15104030	0.650	0.669	0.757	0.619	0.627	0.597	0.610	0.702	0.538	0.554
19	15104040	0.509	0.549	0.586	0.383	0.394	0.284	0.310	0.318	0.165	0.174
20	15104045	0.625	0.643	0.655	0.637	0.598	0.361	0.381	0.369	0.355	0.325
21	15104050	1.192	1.193	1.078	1.083	0.998	1.153	1.171	1.019	1.027	0.936
22	15105010	1.358	1.306	1.251	1.224	1.192	1.605	1.546	1.491	1.450	1.395
23	15105020	1.106	1.032	0.972	0.936	1.008	1.099	1.010	0.935	0.887	0.988
24	20101010	0.871	0.877	0.830	0.927	0.907	0.841	0.846	0.795	0.914	0.894
25	20102010	2.047	1.890	1.736	1.666	1.573	1.943	1.797	1.664	1.591	1.486
26	20103010	0.924	0.930	0.926	0.954	0.915	1.092	1.077	1.055	1.078	1.017
27	20104010	0.633	0.741	0.846	0.720	0.724	0.541	0.672	0.806	0.662	0.674
28	20104020	1.214	1.152	1.091	1.040	0.984	1.094	1.049	0.999	0.928	0.852
29	20105010	0.690	0.697	0.714	0.717	0.705	0.481	0.483	0.489	0.490	0.478
30	20106010	0.726	0.776	0.832	0.790	0.812	0.751	0.791	0.847	0.790	0.819
31	20106015	0.545	0.582	0.632	0.641	0.634	0.329	0.355	0.385	0.393	0.389
32	20106020	0.916	0.858	0.837	0.799	0.795	0.952	0.885	0.863	0.818	0.806
33	20107010	0.661	0.604	0.579	0.729	0.658	0.702	0.617	0.564	0.738	0.648
34	20201010	1.798	1.687	1.684	1.608	1.626	1.789	1.689	1.751	1.665	1.693
35	20201050	1.209	1.300	1.297	1.248	1.233	1.219	1.326	1.345	1.288	1.272

Notes: See Table 3.6 for notes.

Table 3.8: Industry level dependence

ID.	GICS code	AF 2009	AF 2010	AF 2011	AF 2012	AF 2013	AWAF 2009	AWAF 2010	AWAF 2011	AWAF 2012	AWAF 2013
1	10101010	0.287	0.246	0.184	0.159	0.151	0.374	0.360	0.320	0.294	0.289
2	10101020	0.157	0.097	0.079	0.076	0.073	0.293	0.229	0.209	0.207	0.204
3	10102010	0.360	0.293	0.231	0.212	0.194	0.391	0.374	0.355	0.347	0.336
4	10102020	0.037	0.030	0.028	0.026	0.023	0.151	0.138	0.132	0.126	0.118
5	10102030	0.215	0.179	0.149	0.115	0.099	0.333	0.314	0.298	0.261	0.237
6	10102040	0.179	0.115	0.082	0.071	0.061	0.302	0.246	0.211	0.199	0.183
7	10102050	0.349	0.240	0.177	0.157	0.148	0.456	0.381	0.333	0.321	0.313
8	15101010	0.134	0.121	0.102	0.089	0.083	0.256	0.246	0.231	0.215	0.207
9	15101020	0.337	0.294	0.245	0.229	0.215	0.383	0.372	0.359	0.354	0.347
10	15101030	0.248	0.224	0.193	0.178	0.155	0.299	0.293	0.285	0.281	0.264
11	15101040	1.117	0.761	0.508	0.453	0.413	0.792	0.696	0.611	0.587	0.557
12	15101050	0.178	0.152	0.103	0.097	0.088	0.310	0.295	0.247	0.242	0.230
13	15102010	0.309	0.266	0.219	0.205	0.177	0.376	0.363	0.348	0.342	0.313
14	15103010	0.323	0.248	0.213	0.202	0.192	0.404	0.359	0.347	0.342	0.336
15	15103020	0.275	0.246	0.191	0.181	0.147	0.348	0.339	0.308	0.305	0.270
16	15104010	0.781	0.580	0.412	0.370	0.338	0.603	0.558	0.508	0.491	0.470
17	15104020	0.217	0.173	0.149	0.129	0.113	0.331	0.301	0.290	0.269	0.252
18	15104030	0.179	0.157	0.125	0.114	0.102	0.296	0.284	0.261	0.254	0.239
19	15104040	0.875	0.608	0.412	0.298	0.274	0.701	0.618	0.539	0.469	0.448
20	15104045	0.998	0.673	0.442	0.385	0.338	0.731	0.641	0.554	0.526	0.492
21	15104050	0.121	0.105	0.095	0.087	0.083	0.245	0.230	0.224	0.214	0.211
22	15105010	0.454	0.388	0.313	0.292	0.275	0.429	0.417	0.403	0.397	0.389
23	15105020	0.314	0.269	0.221	0.206	0.177	0.385	0.371	0.355	0.348	0.318
24	20101010	0.093	0.082	0.075	0.063	0.059	0.225	0.213	0.206	0.189	0.183
25	20102010	0.076	0.072	0.068	0.066	0.065	0.211	0.205	0.200	0.199	0.197
26	20103010	0.105	0.097	0.088	0.081	0.078	0.227	0.220	0.214	0.206	0.203
27	20104010	0.101	0.082	0.068	0.065	0.060	0.233	0.210	0.193	0.190	0.183
28	20104020	0.603	0.480	0.361	0.324	0.294	0.526	0.498	0.465	0.450	0.432
29	20105010	0.637	0.479	0.349	0.318	0.293	0.602	0.548	0.496	0.480	0.460
30	20106010	0.274	0.204	0.156	0.147	0.128	0.346	0.311	0.282	0.278	0.259
31	20106015	0.907	0.629	0.431	0.386	0.351	0.718	0.632	0.558	0.536	0.509
32	20106020	0.068	0.063	0.056	0.054	0.049	0.193	0.186	0.177	0.175	0.167
33	20107010	0.331	0.267	0.209	0.161	0.149	0.367	0.350	0.332	0.287	0.280
34	20201010	0.143	0.135	0.116	0.112	0.102	0.272	0.266	0.249	0.247	0.235
35	20201050	0.112	0.092	0.081	0.078	0.073	0.237	0.217	0.205	0.204	0.196

Notes: See Table 3.6 for the corresponding rates. For all the industries, see Table B.3 in Appendix B.2.

Table 3.9: Root mean square deviation (RMSD)

ID.	GICS code	AF	AWAF	MF	AWMF
1	10101010	0.014	0.033	0.118	0.207
2	10101020	0.278	0.340	0.410	0.495
3	10102010	0.290	0.307	0.269	0.283
4	10102020	0.192	0.262	0.105	0.123
5	10102030	0.483	0.584	0.624	0.736
6	10102040	0.184	0.227	0.500	0.571
7	10102050	0.629	0.745	0.773	0.856
8	15101010	0.188	0.498	0.154	0.182
9	15101020	0.149	0.290	0.210	0.310
10	15101030	0.420	0.726	0.513	0.787
11	15101040	0.177	0.318	0.230	0.427
12	15101050	0.612	0.662	0.748	0.789
13	15102010	0.059	0.133	0.048	0.063
14	15103010	0.253	0.253	0.246	0.183
15	15103020	0.687	1.063	0.399	0.561
16	15104010	0.103	0.198	0.097	0.217
17	15104020	0.049	0.043	0.100	0.113
18	15104030	0.307	0.352	0.339	0.404
19	15104040	0.494	0.688	0.522	0.753
20	15104045	0.332	0.540	0.369	0.642
21	15104050	0.072	0.069	0.132	0.108
22	15105010	0.247	0.555	0.273	0.503
23	15105020	0.063	0.105	0.059	0.074
24	20101010	0.020	0.024	0.122	0.148
25	20102010	0.726	0.705	0.800	0.714
26	20103010	0.012	0.165	0.071	0.069
27	20104010	0.201	0.210	0.276	0.340
28	20104020	0.018	0.065	0.126	0.088
29	20105010	0.244	0.419	0.296	0.516
30	20106010	0.130	0.069	0.216	0.203
31	20106015	0.303	0.483	0.395	0.630
32	20106020	0.138	0.075	0.165	0.145
33	20107010	0.282	0.256	0.358	0.352
34	20201010	0.799	0.954	0.684	0.719
35	20201050	0.332	0.455	0.26	0.293

Notes: These figures are generated using the rates in Tables 3.6 and 3.7. See Table B.4 in Appendix B.3 for all the industries.

The bias correction is performed in two ways. First, we estimate the potential effects of informative firm censoring in terms of time-varying weights by using firm specific and macroeconomic covariates. It is assumed that all the industries have the same reasons for firm censoring. Second, we adjust the weights for industry level unobserved factors in addition to the firm specific and macroeconomic factors. This is done as reasons for censoring of firms may differ across the industries.

In the empirical analysis, we use a panel of 8471 listed firms on the NYSE, NASDAQ, and NYSE MKT LLC over the period 1980-2013. We compare in-sample and out-of-sample performances of our bias-corrected models with those of Chava et al. (2011) and Atsu and Costantini (2015) in order to give insight into the estimation of default rates during distress market periods. We employ variables such as the time to event, distance to default probability, one year trailing stock return, one year trailing S&P 500 return, 3 month T-bill rate, and a distress indicator at industry level. The results show that the distance to default probability and S&P 500 return covariates are positive and highly significant, while stock return is negatively statistically significant for both the benchmark and bias-corrected models, and the 3 month T-bill rate is negative. Further, the hazard rate adjustment factor, which tends to be higher during distressed markets periods, is more appropriately measured in our bias-corrected models than in the benchmark ones, and the effect of industry level unobserved factors on hazard rates is also better accounted for in our models. Lastly, our bias-corrected models generate smaller standard errors than the benchmarks, and better cope with potential sharp changes in covariates.

We conclude that, when estimating hazard rates, it is important to explicitly account for market events through which firms are censored.

Appendix B

B.1 Industries

Table B.1: Sub-industry codes and names

ID.	GICS code	Name
1	10101010	Oil & Gas Drilling
2	10101020	Oil & Gas Equipment & Services
3	10102010	Integrated Oil & Gas
4	10102020	Oil & Gas Exploration& Production
5	10102030	Oil & Gas Refining& Marketing
6	10102040	Oil & Gas Storage & Transportation
7	10102050	Coal & Consumable Fuels
8	15101010	Commodity Chemicals
9	15101020	Diversified Chemicals
10	15101030	Fertilizers & Agricultural Chemicals
11	15101040	Industrial Gases
12	15101050	Specialty Chemicals
13	15102010	Construction Materials
14	15103010	Metal & Glass Containers
15	15103020	Paper Packaging
16	15104010	Aluminium
17	15104020	Diversified Metals & Mining
18	15104030	Gold
19	15104040	Precious Metals & Minerals

Continued on next page

Table B.1 – *Continued from previous page*

ID.	GICS code	Name
20	15104045	Silver
21	15104050	Steel
22	15105010	Forest Products
23	15105020	Paper Products
24	20101010	Aerospace & Defence
25	20102010	Building Products
26	20103010	Construction & Engineering
27	20104010	Electrical Components & Equipment
28	20104020	Heavy Electrical Equipment
29	20105010	Industrial Conglomerates
30	20106010	Construction Machinery & Heavy Trucks
31	20106015	Agricultural & Farm Machinery
32	20106020	Industrial Machinery
33	20107010	Trading Companies & Distributors
34	20201010	Commercial Printing
35	20201050	Environmental & Facilities Services
36	20201060	Office Services & Supplies
37	20201070	Diversified Support Services
38	20201080	Security & Alarm Services
39	20202010	Human Resource & Employment Services
40	20202020	Research & Consulting Services
41	20301010	Air Freight & Logistics
42	20302010	Airlines
43	20303010	Marine
44	20304010	Railroads
45	20304020	Trucking
46	20305010	Airport Services
47	20305020	Highways & Rail tracks
48	20305030	Marine Ports & Services
49	25101010	Auto Parts & Equipment
50	25101020	Tires & Rubber
51	25102010	Automobile Manufacturers
52	25102020	Motorcycle Manufacturers

Continued on next page

Table B.1 – *Continued from previous page*

ID.	GICS code	Name
53	25201010	Consumer Electronics
54	25201020	Home Furnishings
55	25201030	Homebuilding
56	25201040	Household Appliances
57	25201050	Housewares & Specialties
58	25202010	Leisure Products
59	25202020	Photographic Products
60	25203010	Apparel, Accessories & Luxury Goods
61	25203020	Footwear
62	25203030	Textiles
63	25301010	Casinos & Gaming
64	25301020	Hotels, Resorts & Cruise Lines
65	25301030	Leisure Facilities
66	25301040	Restaurants
67	25302010	Education Services
68	25302020	Specialized Consumer Services
69	25401010	Advertising
70	25401020	Broadcasting
71	25401025	Cable & Satellite
72	25401030	Movies & Entertainment
73	25401040	Publishing
74	25501010	Distributors
75	25502010	Catalogue Retail
76	25502020	Internet Retail
77	25503010	Department Stores
78	25503020	General Merchandise Stores
79	25504010	Apparel Retail
80	25504020	Computer & Electronics Retail
81	25504030	Home Improvement Retail
82	25504040	Specialty Stores
83	25504050	Automotive Retail
84	25504060	Home furnishing Retail
85	30101010	Drug Retail

Continued on next page

Table B.1 – *Continued from previous page*

ID.	GICS code	Name
86	30101020	Food Distributors
87	30101030	Food Retail
88	30101040	Hypermarkets & Super Centers
89	30201010	Brewers
90	30201020	Distillers & Vintners
91	30201030	Soft Drinks
92	30202010	Agricultural Products
93	30202030	Packaged Foods & Meats
94	30203010	Tobacco
95	30301010	Household Products
96	30302010	Personal Products
97	35101010	Health Care Equipment
98	35101020	Health Care Supplies
99	35102010	Health Care Distributors
100	35102015	Health Care Services
101	35102020	Health Care Facilities
102	35102030	Managed Health Care
103	35103010	Health Care Technology
104	35201010	Biotechnology
105	35202010	Pharmaceuticals
106	35203010	Life Sciences Tools & Services
107	40402040	Office REITs
108	45101010	Internet Software & Services
109	45102010	IT Consulting & Other Services
110	45102020	Data Processing & Outsourced Services
111	45103010	Application Software
112	45103020	Systems Software
113	45103030	Home Entertainment Software
114	45201020	Communications Equipment
115	45202010	Computer Hardware
116	45202020	Computer Storage & Peripherals
117	45202030	Technology Hardware, Storage & Peripherals
118	45203010	Electronic Equipment & Instruments

Continued on next page

Table B.1 – *Continued from previous page*

ID.	GICS code	Name
119	45203015	Electronic Components
120	45203020	Electronic Manufacturing Services
121	45203030	Technology Distributors
122	45204010	Office Electronics
123	45301010	Semiconductor Equipment
124	45301020	Semiconductors
125	50101010	Alternative Carriers
126	50101020	Integrated Telecommunication Services
127	50102010	Wireless Telecommunication Services

B.2 Industry level failure rates

Table B.2: Industry level failure rates: AF and AWAF

		AF					AWAF				
ID.	GICS code	2009	2010	2011	2012	2013	2009	2010	2011	2012	2013
1	10101010	0.977	0.910	0.975	1.043	0.986	0.943	0.865	0.931	1.042	0.967
2	10101020	0.482	0.741	0.811	0.771	0.722	0.400	0.687	0.757	0.709	0.660
3	10102010	0.949	0.850	0.797	0.764	0.710	1.117	0.943	0.836	0.776	0.693
4	10102020	1.345	1.365	1.142	1.156	1.192	1.514	1.510	1.221	1.228	1.262
5	10102030	0.448	0.461	0.510	0.442	0.517	0.313	0.341	0.386	0.327	0.416
6	10102040	0.338	0.531	0.748	0.755	0.816	0.264	0.457	0.692	0.694	0.773
7	10102050	0.214	0.293	0.410	0.369	0.371	0.123	0.192	0.290	0.250	0.255
8	15101010	0.922	0.921	1.002	1.137	1.188	1.229	1.174	1.257	1.451	1.498
9	15101020	1.548	1.387	1.252	1.204	1.149	1.877	1.655	1.469	1.390	1.290
10	15101030	2.211	1.843	1.539	1.367	1.420	2.910	2.444	1.965	1.689	1.726
11	15101040	0.769	0.796	0.825	0.828	0.823	0.621	0.651	0.681	0.684	0.682
12	15101050	0.241	0.273	0.393	0.366	0.388	0.198	0.224	0.347	0.310	0.338
13	15102010	1.224	1.106	1.017	0.980	1.059	1.393	1.219	1.086	1.025	1.133
14	15103010	1.379	1.498	1.349	1.304	1.253	1.351	1.531	1.379	1.327	1.253
15	15103020	1.653	1.464	1.500	1.426	1.687	1.972	1.722	1.809	1.701	2.063

Continued on next page

Table B.2 – *Continued from previous page*

		AF					AWAF				
ID.	GICS code	2009	2010	2011	2012	2013	2009	2010	2011	2012	2013
16	15104010	1.016	0.973	0.942	0.923	0.897	1.002	0.939	0.872	0.840	0.802
17	15104020	1.334	1.316	1.104	1.095	1.049	1.242	1.250	0.992	1.005	0.957
18	15104030	0.792	0.772	0.832	0.690	0.693	0.792	0.760	0.827	0.639	0.648
19	15104040	0.601	0.634	0.666	0.503	0.506	0.411	0.440	0.457	0.301	0.312
20	15104045	0.691	0.707	0.719	0.703	0.668	0.479	0.502	0.503	0.486	0.460
21	15104050	1.421	1.354	1.172	1.169	1.072	1.508	1.440	1.190	1.185	1.069
22	15105010	1.591	1.447	1.323	1.285	1.247	2.117	1.896	1.713	1.644	1.555
23	15105020	1.275	1.133	1.032	0.988	1.063	1.399	1.205	1.065	0.999	1.105
24	20101010	1.041	0.998	0.909	1.005	0.980	1.102	1.042	0.933	1.058	1.024
25	20102010	2.533	2.215	1.940	1.842	1.726	2.587	2.240	1.962	1.851	1.705
26	20103010	1.115	1.065	1.014	1.036	0.988	1.441	1.332	1.238	1.249	1.165
27	20104010	0.808	0.876	0.943	0.803	0.799	0.759	0.862	0.967	0.790	0.790
28	20104020	1.370	1.239	1.135	1.077	1.018	1.346	1.214	1.105	1.020	0.935
29	20105010	0.772	0.763	0.771	0.769	0.756	0.602	0.592	0.596	0.592	0.581
30	20106010	0.843	0.867	0.900	0.851	0.870	0.961	0.958	0.981	0.906	0.931
31	20106015	0.626	0.658	0.701	0.705	0.697	0.450	0.479	0.517	0.519	0.517
32	20106020	1.097	0.978	0.918	0.870	0.862	1.250	1.091	1.014	0.948	0.925
33	20107010	0.769	0.687	0.652	0.789	0.718	0.897	0.751	0.667	0.849	0.744
34	20201010	2.229	1.970	1.886	1.778	1.799	2.391	2.106	2.072	1.940	1.954
35	20201050	1.444	1.484	1.420	1.353	1.332	1.597	1.637	1.577	1.489	1.455
36	20201060	1.861	1.653	1.470	1.403	1.323	2.156	1.902	1.687	1.599	1.481
37	20201070	0.070	0.108	0.168	0.180	0.237	0.029	0.047	0.079	0.084	0.141
38	20201080	0.723	1.496	1.329	1.282	1.236	0.548	1.892	1.643	1.553	1.450
39	20202010	1.751	1.654	1.476	1.401	1.403	1.961	1.847	1.649	1.554	1.566
40	20202020	2.573	2.353	2.133	2.054	2.001	3.077	2.788	2.523	2.418	2.323
41	20301010	1.294	1.131	1.193	1.139	1.075	1.661	1.395	1.483	1.388	1.273
42	20302010	1.870	1.799	1.733	1.647	1.539	1.676	1.662	1.656	1.565	1.436
43	20303010	1.058	1.107	1.149	1.032	0.901	1.477	1.450	1.479	1.253	1.021
44	20304010	0.541	0.579	0.626	0.634	0.630	0.383	0.414	0.449	0.455	0.455
45	20304020	1.247	1.110	1.011	1.050	1.142	1.110	0.974	0.878	0.942	1.075
46	20305010	0.829	0.841	0.860	0.858	0.842	0.699	0.713	0.732	0.727	0.709
47	20305020	1.496	1.356	1.247	1.222	1.206	2.191	1.898	1.667	1.609	1.545

Continued on next page

Table B.2 – *Continued from previous page*

		AF					AWAF				
ID.	GICS code	2009	2010	2011	2012	2013	2009	2010	2011	2012	2013
48	20305030	0.888	0.890	0.886	0.879	0.866	0.765	0.768	0.737	0.723	0.711
49	25101010	0.985	0.900	0.834	0.800	0.752	1.083	0.963	0.868	0.817	0.751
50	25101020	2.344	1.965	1.668	1.588	1.516	2.888	2.437	2.090	1.966	1.815
51	25102010	0.681	0.684	0.701	0.700	0.689	0.593	0.577	0.578	0.570	0.556
52	25102020	0.907	0.909	0.919	0.914	0.902	0.807	0.811	0.823	0.812	0.794
53	25201010	1.639	1.456	1.311	1.249	1.155	2.046	1.789	1.593	1.488	1.320
54	25201020	1.411	1.256	1.142	1.234	1.165	1.683	1.461	1.299	1.447	1.328
55	25201030	2.079	1.832	1.679	1.611	1.542	2.403	2.029	1.874	1.785	1.688
56	25201040	0.895	0.861	0.843	0.835	0.819	0.883	0.821	0.781	0.763	0.737
57	25201050	2.206	1.864	1.636	1.551	1.620	2.477	2.063	1.821	1.710	1.792
58	25202010	1.425	1.517	1.351	1.391	1.506	1.411	1.541	1.369	1.435	1.572
59	25202020	3.303	2.822	2.354	2.269	2.201	4.301	3.791	3.295	3.232	3.101
60	25203010	1.741	1.513	1.340	1.327	1.344	1.924	1.656	1.459	1.446	1.452
61	25203020	0.429	0.463	0.508	0.731	0.716	0.285	0.309	0.337	0.680	0.657
62	25203030	3.875	3.364	2.852	2.702	2.581	4.127	3.685	3.273	3.114	2.943
63	25301010	1.128	1.002	0.916	0.872	0.889	1.240	1.077	0.961	0.901	0.919
64	25301020	2.025	1.788	1.846	2.003	2.003	2.067	1.820	1.939	2.172	2.163
65	25301030	0.179	0.284	0.363	0.380	0.393	0.086	0.169	0.223	0.233	0.248
66	25301040	1.692	1.619	1.432	1.415	1.365	2.304	2.150	1.887	1.850	1.750
67	25302010	0.623	0.623	0.769	0.843	1.003	0.872	0.777	0.945	1.025	1.227
68	25302020	0.821	0.777	0.765	0.751	1.020	0.911	0.801	0.740	0.709	1.125
69	25401010	0.255	0.308	0.378	0.387	0.395	0.195	0.236	0.294	0.296	0.306
70	25401020	2.078	1.865	1.757	1.726	1.685	2.318	2.064	1.951	1.915	1.848
71	25401025	1.037	1.210	1.123	1.272	1.378	0.956	1.211	1.131	1.351	1.495
72	25401030	1.696	1.763	1.689	1.610	1.509	1.934	2.022	1.945	1.842	1.695
73	25401040	1.082	0.966	0.989	0.946	0.982	1.131	0.981	1.011	0.953	0.992
74	25501010	3.238	2.896	2.606	2.529	2.424	3.529	3.138	2.855	2.771	2.630
75	25502010	1.767	1.529	1.342	1.285	1.223	1.866	1.604	1.400	1.325	1.229
76	25502020	0.857	0.818	0.809	0.762	0.812	0.694	0.669	0.668	0.614	0.706
77	25503010	2.421	2.046	1.781	1.692	1.822	2.630	2.168	1.928	1.824	2.012
78	25503020	1.162	1.144	1.131	1.118	1.097	1.332	1.282	1.254	1.229	1.189
79	25504010	1.058	0.927	0.842	1.000	0.926	1.029	0.886	0.789	0.975	0.889

Continued on next page

Table B.2 – *Continued from previous page*

		AF					AWAF				
ID.	GICS code	2009	2010	2011	2012	2013	2009	2010	2011	2012	2013
80	25504020	1.143	1.061	1.007	0.976	0.940	1.178	1.065	0.989	0.938	0.883
81	25504030	2.163	1.827	1.568	1.497	1.433	2.643	2.203	1.877	1.762	1.630
82	25504040	2.255	1.958	1.812	1.862	2.030	2.342	2.024	1.887	1.953	2.158
83	25504050	0.755	0.725	0.874	0.836	0.781	0.946	0.832	1.004	0.925	0.825
84	25504060	0.829	0.795	0.803	0.970	0.933	0.821	0.733	0.723	0.989	0.927
85	30101010	1.900	1.663	1.457	1.378	1.308	2.681	2.305	1.980	1.819	1.660
86	30101020	0.813	0.803	0.804	0.798	0.922	0.645	0.634	0.632	0.624	0.841
87	30101030	1.044	1.020	0.997	0.982	1.020	1.324	1.248	1.189	1.154	1.185
88	30101040	0.266	0.330	0.416	0.402	0.410	0.149	0.192	0.250	0.237	0.249
89	30201010	0.665	0.688	0.721	0.724	0.714	0.491	0.511	0.538	0.539	0.535
90	30201020	1.341	1.207	1.106	1.061	1.009	1.438	1.270	1.140	1.071	0.991
91	30201030	1.290	1.139	1.040	0.993	0.933	1.344	1.153	1.025	0.958	0.872
92	30202010	2.268	1.916	1.635	1.320	1.235	2.701	2.275	1.938	1.390	1.272
93	30202030	1.012	0.987	0.960	1.009	1.075	1.255	1.191	1.139	1.189	1.252
94	30203010	1.041	0.972	0.931	0.913	0.883	1.182	1.058	0.973	0.938	0.884
95	30301010	1.119	1.017	0.890	0.867	0.833	1.204	1.056	0.866	0.828	0.779
96	30302010	1.144	1.139	1.130	1.172	1.100	1.430	1.388	1.351	1.392	1.279
97	35101010	1.575	1.520	1.391	1.418	1.331	1.857	1.765	1.606	1.636	1.518
98	35101020	1.035	0.950	0.888	0.912	0.840	1.141	1.014	0.930	0.951	0.855
99	35102010	1.884	1.700	1.685	1.769	1.703	2.666	2.386	2.367	2.484	2.359
100	35102015	0.184	0.345	0.537	0.637	0.629	0.142	0.294	0.489	0.597	0.593
101	35102020	1.498	1.533	1.644	1.729	1.805	1.443	1.528	1.712	1.822	1.921
102	35102030	1.248	1.124	1.047	1.019	1.120	1.457	1.289	1.184	1.136	1.264
103	35103010	0.483	0.726	0.912	0.872	0.828	0.376	0.672	0.895	0.841	0.785
104	35201010	0.994	0.946	0.858	0.869	0.860	1.001	0.957	0.878	0.892	0.885
105	35202010	0.839	0.932	0.892	0.908	0.935	0.991	1.069	1.014	1.004	1.027
106	35203010	0.244	0.291	0.399	0.548	0.521	0.174	0.206	0.308	0.507	0.476
107	40402040	0.996	0.996	0.997	0.994	0.994	0.988	0.991	0.992	0.984	0.985
108	45101010	1.009	1.094	1.165	1.092	1.019	0.998	1.079	1.177	1.099	1.032
109	45102010	1.486	1.463	1.648	1.608	1.717	1.620	1.580	1.781	1.732	1.849
110	45102020	0.633	0.687	0.682	0.806	0.808	0.693	0.732	0.707	0.867	0.861
111	45103010	1.266	1.205	1.217	1.247	1.209	1.299	1.229	1.244	1.274	1.233

Continued on next page

Table B.2 – *Continued from previous page*

		AF					AWAF				
ID.	GICS code	2009	2010	2011	2012	2013	2009	2010	2011	2012	2013
112	45103020	1.134	1.246	1.215	1.283	1.319	1.354	1.450	1.409	1.494	1.519
113	45103030	0.808	0.597	0.726	0.700	0.668	0.672	0.457	0.617	0.579	0.543
114	45201020	0.654	0.735	0.862	0.876	0.953	0.584	0.676	0.823	0.839	0.928
115	45202010	2.243	2.053	1.853	1.807	1.960	2.320	2.150	1.991	1.956	2.169
116	45202020	1.721	1.850	1.872	1.962	2.055	2.108	2.258	2.302	2.433	2.560
117	45202030	0.177	0.221	0.273	0.276	0.274	0.086	0.114	0.145	0.146	0.150
118	45203010	0.952	0.989	0.991	1.056	0.987	1.057	1.087	1.079	1.153	1.064
119	45203015	0.410	0.434	0.471	0.654	0.809	0.262	0.277	0.299	0.553	0.777
120	45203020	0.587	0.572	0.644	0.617	0.637	0.514	0.491	0.571	0.537	0.572
121	45203030	0.503	0.633	0.636	0.833	0.800	0.598	0.720	0.674	0.947	0.889
122	45204010	1.217	1.155	1.105	1.098	1.092	1.369	1.273	1.197	1.187	1.170
123	45301010	0.492	0.474	0.469	0.547	0.554	0.462	0.427	0.403	0.502	0.513
124	45301020	0.874	0.830	0.809	0.810	0.842	0.908	0.859	0.841	0.844	0.876
125	50101010	0.231	0.357	0.597	0.663	0.686	0.179	0.286	0.529	0.593	0.630
126	50101020	1.008	1.070	1.116	1.057	1.034	1.117	1.165	1.220	1.143	1.106
127	50102010	2.321	2.248	2.045	1.944	1.912	2.851	2.757	2.549	2.419	2.354

Table B.3: Industry level failure rates: MF and AWMF

		MF					AWMF				
ID.	GICS code	2009	2010	2011	2012	2013	2009	2010	2011	2012	2013
1	10101010	0.846	0.821	0.910	0.980	0.929	0.739	0.717	0.806	0.913	0.855
2	10101020	0.365	0.621	0.723	0.691	0.651	0.282	0.535	0.629	0.596	0.561
3	10102010	0.831	0.769	0.735	0.705	0.650	0.882	0.788	0.724	0.676	0.601
4	10102020	1.105	1.181	1.027	1.049	1.088	1.142	1.212	1.027	1.047	1.088
5	10102030	0.317	0.354	0.419	0.357	0.440	0.205	0.244	0.294	0.250	0.335
6	10102040	0.229	0.417	0.646	0.660	0.727	0.172	0.338	0.559	0.568	0.647
7	10102050	0.112	0.188	0.307	0.270	0.276	0.060	0.117	0.204	0.171	0.176
8	15101010	0.759	0.799	0.912	1.045	1.098	0.931	0.947	1.070	1.247	1.303
9	15101020	1.314	1.247	1.178	1.142	1.095	1.437	1.362	1.284	1.232	1.159
10	15101030	1.782	1.594	1.413	1.275	1.321	2.142	1.937	1.669	1.462	1.506

Continued on next page

Table B.3 – *Continued from previous page*

		MF					AWMF				
ID.	GICS code	2009	2010	2011	2012	2013	2009	2010	2011	2012	2013
11	15101040	0.725	0.756	0.789	0.795	0.792	0.532	0.562	0.587	0.596	0.593
12	15101050	0.159	0.192	0.316	0.292	0.318	0.128	0.157	0.269	0.241	0.269
13	15102010	1.064	1.007	0.957	0.927	1.005	1.093	1.019	0.953	0.909	1.012
14	15103010	1.191	1.336	1.259	1.228	1.185	1.067	1.269	1.208	1.177	1.125
15	15103020	1.395	1.308	1.383	1.329	1.538	1.507	1.412	1.552	1.481	1.782
16	15104010	0.940	0.925	0.909	0.892	0.867	0.845	0.827	0.785	0.759	0.722
17	15104020	1.138	1.172	1.026	1.022	0.981	0.966	1.030	0.855	0.875	0.840
18	15104030	0.650	0.669	0.757	0.619	0.627	0.597	0.610	0.702	0.538	0.554
19	15104040	0.509	0.549	0.586	0.383	0.394	0.284	0.310	0.318	0.165	0.174
20	15104045	0.625	0.643	0.655	0.637	0.598	0.361	0.381	0.369	0.355	0.325
21	15104050	1.192	1.193	1.078	1.083	0.998	1.153	1.171	1.019	1.027	0.936
22	15105010	1.358	1.306	1.251	1.224	1.192	1.605	1.546	1.491	1.450	1.395
23	15105020	1.106	1.032	0.972	0.936	1.008	1.099	1.010	0.935	0.887	0.988
24	20101010	0.871	0.877	0.830	0.927	0.907	0.841	0.846	0.795	0.914	0.894
25	20102010	2.047	1.890	1.736	1.666	1.573	1.943	1.797	1.664	1.591	1.486
26	20103010	0.924	0.930	0.926	0.954	0.915	1.092	1.077	1.055	1.078	1.017
27	20104010	0.633	0.741	0.846	0.720	0.724	0.541	0.672	0.806	0.662	0.674
28	20104020	1.214	1.152	1.091	1.040	0.984	1.094	1.049	0.999	0.928	0.852
29	20105010	0.690	0.697	0.714	0.717	0.705	0.481	0.483	0.489	0.490	0.478
30	20106010	0.726	0.776	0.832	0.790	0.812	0.751	0.791	0.847	0.790	0.819
31	20106015	0.545	0.582	0.632	0.641	0.634	0.329	0.355	0.385	0.393	0.389
32	20106020	0.916	0.858	0.837	0.799	0.795	0.952	0.885	0.863	0.818	0.806
33	20107010	0.661	0.604	0.579	0.729	0.658	0.702	0.617	0.564	0.738	0.648
34	20201010	1.798	1.687	1.684	1.608	1.626	1.789	1.689	1.751	1.665	1.693
35	20201050	1.209	1.300	1.297	1.248	1.233	1.219	1.326	1.345	1.288	1.272
36	20201060	1.540	1.446	1.348	1.300	1.233	1.632	1.539	1.443	1.388	1.302
37	20201070	0.019	0.035	0.069	0.078	0.142	0.006	0.011	0.022	0.025	0.078
38	20201080	0.667	1.363	1.270	1.234	1.194	0.443	1.560	1.456	1.397	1.327
39	20202010	1.450	1.440	1.348	1.293	1.297	1.485	1.490	1.405	1.343	1.367
40	20202020	2.077	2.002	1.898	1.845	1.805	2.295	2.215	2.117	2.056	1.998
41	20301010	1.121	1.028	1.119	1.077	1.019	1.286	1.156	1.285	1.220	1.133
42	20302010	1.548	1.560	1.565	1.504	1.417	1.290	1.357	1.420	1.362	1.267

Continued on next page

Table B.3 – *Continued from previous page*

		MF					AWMF				
ID.	GICS code	2009	2010	2011	2012	2013	2009	2010	2011	2012	2013
43	20303010	0.947	1.016	1.084	0.979	0.850	1.164	1.204	1.284	1.105	0.907
44	20304010	0.441	0.485	0.538	0.552	0.551	0.258	0.284	0.310	0.321	0.320
45	20304020	1.057	0.989	0.936	0.979	1.067	0.859	0.801	0.755	0.821	0.946
46	20305010	0.798	0.812	0.832	0.831	0.815	0.629	0.640	0.653	0.651	0.626
47	20305020	1.348	1.282	1.218	1.199	1.183	1.672	1.587	1.510	1.473	1.438
48	20305030	0.870	0.870	0.864	0.856	0.844	0.713	0.710	0.658	0.645	0.628
49	25101010	0.836	0.799	0.765	0.737	0.692	0.835	0.789	0.743	0.707	0.654
50	25101020	1.834	1.672	1.524	1.469	1.414	2.064	1.904	1.771	1.696	1.604
51	25102010	0.592	0.607	0.632	0.635	0.625	0.467	0.467	0.472	0.469	0.454
52	25102020	0.893	0.894	0.904	0.899	0.887	0.766	0.765	0.772	0.760	0.736
53	25201010	1.383	1.300	1.225	1.177	1.095	1.557	1.459	1.377	1.306	1.175
54	25201020	1.203	1.130	1.071	1.159	1.100	1.295	1.204	1.131	1.267	1.179
55	25201030	1.687	1.577	1.515	1.468	1.415	1.799	1.627	1.588	1.534	1.469
56	25201040	0.795	0.790	0.790	0.787	0.772	0.714	0.696	0.680	0.670	0.649
57	25201050	1.786	1.610	1.490	1.429	1.485	1.855	1.662	1.559	1.486	1.566
58	25202010	1.202	1.332	1.243	1.286	1.387	1.087	1.256	1.177	1.246	1.374
59	25202020	2.356	2.166	1.957	1.904	1.856	2.829	2.675	2.510	2.471	2.410
60	25203010	1.442	1.323	1.226	1.224	1.242	1.459	1.340	1.245	1.249	1.267
61	25203020	0.313	0.349	0.395	0.671	0.658	0.166	0.182	0.200	0.585	0.565
62	25203030	2.779	2.556	2.306	2.213	2.131	2.883	2.745	2.595	2.501	2.405
63	25301010	0.956	0.891	0.843	0.808	0.828	0.953	0.881	0.825	0.782	0.806
64	25301020	1.658	1.549	1.655	1.788	1.791	1.564	1.472	1.644	1.851	1.863
65	25301030	0.075	0.170	0.243	0.262	0.279	0.026	0.091	0.128	0.136	0.148
66	25301040	1.402	1.410	1.304	1.300	1.259	1.738	1.727	1.599	1.588	1.520
67	25302010	0.516	0.531	0.704	0.786	0.947	0.685	0.642	0.823	0.901	1.086
68	25302020	0.742	0.712	0.707	0.695	0.978	0.749	0.683	0.642	0.616	1.016
69	25401010	0.167	0.218	0.289	0.299	0.311	0.126	0.163	0.216	0.221	0.231
70	25401020	1.694	1.605	1.581	1.565	1.535	1.741	1.653	1.648	1.639	1.600
71	25401025	0.924	1.106	1.065	1.203	1.292	0.766	1.015	0.997	1.194	1.325
72	25401030	1.411	1.530	1.528	1.472	1.389	1.469	1.628	1.649	1.584	1.478
73	25401040	0.927	0.865	0.917	0.883	0.920	0.878	0.808	0.873	0.831	0.874
74	25501010	2.559	2.405	2.259	2.208	2.127	2.610	2.470	2.374	2.332	2.241

Continued on next page

Table B.3 – *Continued from previous page*

		MF					AWMF				
ID.	GICS code	2009	2010	2011	2012	2013	2009	2010	2011	2012	2013
75	25502010	1.490	1.374	1.268	1.224	1.171	1.450	1.342	1.244	1.193	1.120
76	25502020	0.753	0.741	0.750	0.703	0.760	0.542	0.547	0.565	0.517	0.612
77	25503010	1.917	1.738	1.602	1.541	1.636	1.949	1.740	1.646	1.581	1.742
78	25503020	0.976	1.013	1.048	1.045	1.032	1.025	1.052	1.086	1.078	1.056
79	25504010	0.895	0.822	0.771	0.927	0.861	0.792	0.723	0.672	0.844	0.776
80	25504020	1.024	0.988	0.964	0.937	0.903	0.957	0.916	0.885	0.846	0.797
81	25504030	1.737	1.586	1.451	1.400	1.350	1.929	1.761	1.625	1.552	1.467
82	25504040	1.836	1.686	1.630	1.683	1.826	1.764	1.627	1.600	1.674	1.862
83	25504050	0.672	0.654	0.823	0.787	0.730	0.766	0.705	0.885	0.822	0.732
84	25504060	0.751	0.732	0.752	0.932	0.896	0.673	0.619	0.625	0.890	0.838
85	30101010	1.577	1.469	1.357	1.298	1.240	1.977	1.834	1.690	1.582	1.472
86	30101020	0.734	0.741	0.752	0.750	0.884	0.522	0.527	0.529	0.526	0.756
87	30101030	0.869	0.894	0.913	0.908	0.947	1.008	1.013	1.018	1.001	1.038
88	30101040	0.137	0.195	0.280	0.270	0.284	0.058	0.084	0.121	0.114	0.123
89	30201010	0.594	0.620	0.657	0.664	0.656	0.376	0.392	0.410	0.417	0.410
90	30201020	1.169	1.107	1.051	1.014	0.966	1.136	1.071	1.010	0.960	0.893
91	30201030	1.130	1.047	0.987	0.947	0.888	1.068	0.976	0.906	0.854	0.779
92	30202010	1.814	1.646	1.493	1.235	1.165	1.991	1.814	1.656	1.219	1.132
93	30202030	0.839	0.863	0.876	0.929	0.993	0.952	0.963	0.970	1.025	1.091
94	30203010	0.933	0.902	0.884	0.871	0.843	0.954	0.904	0.865	0.842	0.796
95	30301010	0.987	0.934	0.833	0.815	0.783	0.960	0.893	0.755	0.728	0.686
96	30302010	0.976	1.016	1.046	1.092	1.029	1.101	1.135	1.161	1.209	1.124
97	35101010	1.297	1.318	1.259	1.295	1.221	1.401	1.418	1.359	1.402	1.316
98	35101020	0.880	0.838	0.814	0.843	0.778	0.880	0.825	0.795	0.823	0.746
99	35102010	1.558	1.485	1.528	1.603	1.551	1.992	1.904	1.986	2.097	2.018
100	35102015	0.113	0.260	0.455	0.555	0.555	0.089	0.215	0.393	0.491	0.496
101	35102020	1.177	1.292	1.464	1.549	1.623	1.047	1.196	1.429	1.538	1.644
102	35102030	1.083	1.024	0.986	0.967	1.063	1.137	1.072	1.036	1.006	1.127
103	35103010	0.381	0.642	0.846	0.813	0.772	0.270	0.549	0.771	0.732	0.686
104	35201010	0.813	0.817	0.774	0.791	0.787	0.751	0.766	0.741	0.762	0.765
105	35202010	0.678	0.798	0.802	0.822	0.853	0.739	0.852	0.854	0.853	0.885
106	35203010	0.138	0.181	0.296	0.466	0.444	0.099	0.123	0.218	0.414	0.390

Continued on next page

Table B.3 – *Continued from previous page*

		MF					AWMF				
ID.	GICS code	2009	2010	2011	2012	2013	2009	2010	2011	2012	2013
107	40402040	0.995	0.996	0.997	0.993	0.993	0.987	0.989	0.990	0.980	0.981
108	45101010	0.809	0.932	1.046	0.989	0.930	0.740	0.856	0.988	0.934	0.889
109	45102010	1.236	1.277	1.489	1.466	1.564	1.230	1.276	1.509	1.487	1.601
110	45102020	0.509	0.586	0.607	0.733	0.740	0.519	0.584	0.594	0.743	0.746
111	45103010	1.044	1.045	1.101	1.138	1.108	0.981	0.988	1.053	1.091	1.068
112	45103020	0.946	1.090	1.109	1.179	1.216	1.028	1.170	1.197	1.284	1.319
113	45103030	0.725	0.502	0.655	0.631	0.600	0.540	0.344	0.511	0.480	0.446
114	45201020	0.510	0.616	0.767	0.787	0.865	0.422	0.528	0.685	0.708	0.795
115	45202010	1.814	1.750	1.658	1.629	1.741	1.747	1.728	1.693	1.682	1.863
116	45202020	1.412	1.588	1.673	1.757	1.837	1.585	1.801	1.932	2.060	2.184
117	45202030	0.076	0.108	0.150	0.154	0.156	0.027	0.040	0.055	0.056	0.059
118	45203010	0.793	0.865	0.902	0.969	0.909	0.804	0.879	0.917	0.992	0.925
119	45203015	0.292	0.316	0.353	0.584	0.755	0.145	0.155	0.166	0.458	0.681
120	45203020	0.478	0.483	0.570	0.546	0.572	0.385	0.386	0.473	0.448	0.485
121	45203030	0.402	0.549	0.559	0.777	0.746	0.463	0.593	0.571	0.830	0.785
122	45204010	1.127	1.106	1.082	1.078	1.074	1.167	1.149	1.127	1.127	1.120
123	45301010	0.398	0.391	0.389	0.478	0.489	0.348	0.334	0.322	0.420	0.433
124	45301020	0.716	0.717	0.730	0.738	0.772	0.682	0.689	0.710	0.722	0.758
125	50101010	0.132	0.239	0.478	0.549	0.587	0.109	0.193	0.405	0.463	0.508
126	50101020	0.847	0.943	1.023	0.978	0.960	0.853	0.945	1.041	0.988	0.967
127	50102010	1.854	1.892	1.808	1.739	1.719	2.106	2.170	2.123	2.044	2.014

B.3 Industry level dependence

Table B.4: Classical AF and bias-corrected AWAF

		AF					AWAF				
ID.	GICS code	2009	2010	2011	2012	2013	2009	2010	2011	2012	2013
1	10101010	0.287	0.246	0.184	0.159	0.151	0.374	0.360	0.320	0.294	0.289
2	10101020	0.157	0.097	0.079	0.076	0.073	0.293	0.229	0.209	0.207	0.204

Continued on next page

Table B.4 – *Continued from previous page*

		AF					AWAF				
ID.	GICS code	2009	2010	2011	2012	2013	2009	2010	2011	2012	2013
3	10102010	0.360	0.293	0.231	0.212	0.194	0.391	0.374	0.355	0.347	0.336
4	10102020	0.037	0.030	0.028	0.026	0.023	0.151	0.138	0.132	0.126	0.118
5	10102030	0.215	0.179	0.149	0.115	0.099	0.333	0.314	0.298	0.261	0.237
6	10102040	0.179	0.115	0.082	0.071	0.061	0.302	0.246	0.211	0.199	0.183
7	10102050	0.349	0.240	0.177	0.157	0.148	0.456	0.381	0.333	0.321	0.313
8	15101010	0.134	0.121	0.102	0.089	0.083	0.256	0.246	0.231	0.215	0.207
9	15101020	0.337	0.294	0.245	0.229	0.215	0.383	0.372	0.359	0.354	0.347
10	15101030	0.248	0.224	0.193	0.178	0.155	0.299	0.293	0.285	0.281	0.264
11	15101040	1.117	0.761	0.508	0.453	0.413	0.792	0.696	0.611	0.587	0.557
12	15101050	0.178	0.152	0.103	0.097	0.088	0.310	0.295	0.247	0.242	0.230
13	15102010	0.309	0.266	0.219	0.205	0.177	0.376	0.363	0.348	0.342	0.313
14	15103010	0.323	0.248	0.213	0.202	0.192	0.404	0.359	0.347	0.342	0.336
15	15103020	0.275	0.246	0.191	0.181	0.147	0.348	0.339	0.308	0.305	0.270
16	15104010	0.781	0.580	0.412	0.370	0.338	0.603	0.558	0.508	0.491	0.470
17	15104020	0.217	0.173	0.149	0.129	0.113	0.331	0.301	0.290	0.269	0.252
18	15104030	0.179	0.157	0.125	0.114	0.102	0.296	0.284	0.261	0.254	0.239
19	15104040	0.875	0.608	0.412	0.298	0.274	0.701	0.618	0.539	0.469	0.448
20	15104045	0.998	0.673	0.442	0.385	0.338	0.731	0.641	0.554	0.526	0.492
21	15104050	0.121	0.105	0.095	0.087	0.083	0.245	0.230	0.224	0.214	0.211
22	15105010	0.454	0.388	0.313	0.292	0.275	0.429	0.417	0.403	0.397	0.389
23	15105020	0.314	0.269	0.221	0.206	0.177	0.385	0.371	0.355	0.348	0.318
24	20101010	0.093	0.082	0.075	0.063	0.059	0.225	0.213	0.206	0.189	0.183
25	20102010	0.076	0.072	0.068	0.066	0.065	0.211	0.205	0.200	0.199	0.197
26	20103010	0.105	0.097	0.088	0.081	0.078	0.227	0.220	0.214	0.206	0.203
27	20104010	0.101	0.082	0.068	0.065	0.060	0.233	0.210	0.193	0.190	0.183
28	20104020	0.603	0.480	0.361	0.324	0.294	0.526	0.498	0.465	0.450	0.432
29	20105010	0.637	0.479	0.349	0.318	0.293	0.602	0.548	0.496	0.480	0.460
30	20106010	0.274	0.204	0.156	0.147	0.128	0.346	0.311	0.282	0.278	0.259
31	20106015	0.907	0.629	0.431	0.386	0.351	0.718	0.632	0.558	0.536	0.509
32	20106020	0.068	0.063	0.056	0.054	0.049	0.193	0.186	0.177	0.175	0.167
33	20107010	0.331	0.267	0.209	0.161	0.149	0.367	0.350	0.332	0.287	0.280
34	20201010	0.143	0.135	0.116	0.112	0.102	0.272	0.266	0.249	0.247	0.235

Continued on next page

Table B.4 – *Continued from previous page*

		AF					AWAF				
ID.	GICS code	2009	2010	2011	2012	2013	2009	2010	2011	2012	2013
35	20201050	0.112	0.092	0.081	0.078	0.073	0.237	0.217	0.205	0.204	0.196
36	20201060	0.139	0.130	0.119	0.115	0.111	0.255	0.249	0.243	0.241	0.239
37	20201070	0.318	0.251	0.195	0.178	0.147	0.471	0.430	0.394	0.381	0.323
38	20201080	1.045	0.555	0.412	0.375	0.346	0.761	0.503	0.476	0.465	0.451
39	20202010	0.125	0.108	0.100	0.097	0.089	0.241	0.227	0.221	0.219	0.209
40	20202020	0.048	0.045	0.043	0.042	0.040	0.167	0.160	0.155	0.153	0.149
41	20301010	0.316	0.269	0.201	0.189	0.178	0.351	0.340	0.305	0.302	0.297
42	20302010	0.175	0.147	0.124	0.120	0.116	0.314	0.288	0.268	0.265	0.262
43	20303010	0.517	0.334	0.234	0.211	0.189	0.424	0.366	0.324	0.317	0.307
44	20304010	0.798	0.563	0.390	0.351	0.321	0.689	0.607	0.536	0.516	0.490
45	20304020	0.162	0.147	0.130	0.115	0.098	0.301	0.292	0.282	0.264	0.238
46	20305010	1.217	0.810	0.531	0.470	0.423	0.826	0.719	0.627	0.600	0.565
47	20305020	1.128	0.806	0.548	0.493	0.455	0.710	0.663	0.608	0.591	0.570
48	20305030	1.324	0.866	0.550	0.483	0.436	0.854	0.740	0.629	0.599	0.565
49	25101010	0.155	0.140	0.123	0.117	0.111	0.273	0.265	0.257	0.254	0.249
50	25101020	0.560	0.472	0.373	0.344	0.320	0.483	0.470	0.451	0.443	0.432
51	25102010	0.587	0.444	0.326	0.297	0.274	0.548	0.505	0.462	0.448	0.431
52	25102020	1.361	0.889	0.574	0.506	0.457	0.873	0.756	0.657	0.627	0.590
53	25201010	0.275	0.245	0.210	0.198	0.185	0.336	0.328	0.319	0.315	0.308
54	25201020	0.262	0.232	0.197	0.170	0.161	0.341	0.331	0.320	0.295	0.291
55	25201030	0.101	0.095	0.089	0.087	0.085	0.226	0.220	0.215	0.213	0.211
56	25201040	0.472	0.379	0.292	0.269	0.252	0.491	0.462	0.431	0.421	0.408
57	25201050	0.209	0.192	0.171	0.164	0.145	0.316	0.309	0.301	0.298	0.279
58	25202010	0.166	0.127	0.116	0.105	0.090	0.290	0.256	0.249	0.237	0.218
59	25202020	0.352	0.329	0.291	0.280	0.271	0.401	0.397	0.392	0.391	0.389
60	25203010	0.092	0.086	0.080	0.074	0.065	0.214	0.208	0.202	0.195	0.184
61	25203020	0.669	0.474	0.331	0.247	0.231	0.643	0.566	0.499	0.391	0.380
62	25203030	0.178	0.172	0.163	0.159	0.156	0.301	0.297	0.293	0.291	0.289
63	25301010	0.159	0.143	0.126	0.120	0.107	0.271	0.263	0.255	0.252	0.238
64	25301020	0.127	0.120	0.097	0.085	0.079	0.255	0.249	0.227	0.210	0.203
65	25301030	0.418	0.284	0.224	0.208	0.197	0.526	0.425	0.396	0.386	0.374
66	25301040	0.077	0.066	0.062	0.058	0.054	0.186	0.174	0.169	0.164	0.160

Continued on next page

Table B.4 – *Continued from previous page*

		AF					AWAF				
ID.	GICS code	2009	2010	2011	2012	2013	2009	2010	2011	2012	2013
67	25302010	0.569	0.422	0.277	0.225	0.173	0.461	0.432	0.367	0.331	0.285
68	25302020	0.665	0.485	0.347	0.313	0.240	0.538	0.498	0.457	0.443	0.360
69	25401010	0.187	0.162	0.137	0.129	0.124	0.316	0.303	0.290	0.285	0.279
70	25401020	0.065	0.061	0.056	0.052	0.050	0.186	0.180	0.172	0.167	0.163
71	25401025	0.512	0.350	0.281	0.237	0.203	0.470	0.397	0.380	0.345	0.316
72	25401030	0.124	0.101	0.089	0.087	0.084	0.246	0.224	0.212	0.211	0.208
73	25401040	0.206	0.182	0.143	0.136	0.121	0.315	0.305	0.278	0.274	0.258
74	25501010	0.067	0.064	0.061	0.061	0.060	0.199	0.193	0.188	0.187	0.185
75	25502010	0.478	0.400	0.316	0.292	0.271	0.456	0.439	0.418	0.409	0.398
76	25502020	0.468	0.371	0.285	0.255	0.210	0.454	0.428	0.402	0.389	0.344
77	25503010	0.254	0.232	0.206	0.196	0.173	0.349	0.340	0.332	0.328	0.304
78	25503020	0.217	0.197	0.174	0.166	0.160	0.337	0.328	0.319	0.316	0.311
79	25504010	0.157	0.141	0.123	0.096	0.093	0.277	0.269	0.259	0.228	0.225
80	25504020	0.540	0.431	0.330	0.300	0.277	0.501	0.474	0.444	0.432	0.417
81	25504030	0.535	0.449	0.355	0.328	0.306	0.494	0.477	0.455	0.446	0.434
82	25504040	0.075	0.071	0.064	0.057	0.049	0.202	0.196	0.187	0.178	0.163
83	25504050	0.628	0.462	0.299	0.270	0.244	0.491	0.460	0.388	0.379	0.365
84	25504060	0.670	0.494	0.361	0.299	0.276	0.552	0.508	0.469	0.412	0.400
85	30101010	0.365	0.321	0.267	0.248	0.233	0.382	0.374	0.363	0.357	0.350
86	30101020	0.660	0.497	0.361	0.327	0.273	0.606	0.553	0.500	0.484	0.420
87	30101030	0.124	0.114	0.103	0.100	0.091	0.251	0.244	0.237	0.235	0.224
88	30101040	0.506	0.386	0.292	0.255	0.238	0.570	0.516	0.469	0.445	0.427
89	30201010	0.961	0.656	0.443	0.395	0.359	0.736	0.644	0.565	0.542	0.514
90	30201020	0.418	0.349	0.278	0.257	0.238	0.431	0.414	0.394	0.386	0.375
91	30201030	0.410	0.339	0.268	0.246	0.227	0.434	0.414	0.392	0.383	0.371
92	30202010	0.305	0.274	0.234	0.204	0.191	0.365	0.356	0.346	0.332	0.326
93	30202030	0.100	0.087	0.075	0.067	0.057	0.219	0.207	0.196	0.185	0.172
94	30203010	0.512	0.408	0.312	0.286	0.265	0.474	0.450	0.424	0.414	0.401
95	30301010	0.385	0.320	0.246	0.228	0.212	0.426	0.407	0.381	0.373	0.363
96	30302010	0.181	0.148	0.121	0.109	0.105	0.276	0.257	0.240	0.230	0.227
97	35101010	0.036	0.031	0.029	0.026	0.025	0.146	0.135	0.128	0.123	0.120
98	35101020	0.177	0.142	0.125	0.111	0.106	0.282	0.262	0.254	0.241	0.237

Continued on next page

Table B.4 – *Continued from previous page*

		AF					AWAF				
ID.	GICS code	2009	2010	2011	2012	2013	2009	2010	2011	2012	2013
99	35102010	0.155	0.145	0.123	0.112	0.109	0.253	0.247	0.233	0.224	0.222
100	35102015	0.194	0.112	0.076	0.062	0.060	0.322	0.246	0.205	0.185	0.183
101	35102020	0.083	0.074	0.063	0.057	0.052	0.215	0.201	0.187	0.178	0.169
102	35102030	0.312	0.268	0.223	0.209	0.182	0.358	0.347	0.335	0.331	0.306
103	35103010	0.483	0.274	0.179	0.168	0.158	0.503	0.370	0.305	0.300	0.294
104	35201010	0.047	0.040	0.036	0.032	0.029	0.161	0.149	0.141	0.134	0.127
105	35202010	0.062	0.049	0.045	0.038	0.034	0.180	0.162	0.155	0.145	0.137
106	35203010	0.364	0.285	0.197	0.142	0.133	0.453	0.420	0.356	0.283	0.276
107	40402040	1.547	1.003	0.637	0.561	0.514	0.956	0.828	0.715	0.684	0.649
108	45101010	0.056	0.046	0.039	0.037	0.035	0.169	0.155	0.143	0.140	0.136
109	45102010	0.091	0.076	0.058	0.054	0.046	0.211	0.195	0.173	0.168	0.156
110	45102020	0.147	0.122	0.108	0.087	0.080	0.272	0.253	0.245	0.218	0.209
111	45103010	0.040	0.035	0.030	0.028	0.026	0.153	0.143	0.133	0.127	0.123
112	45103020	0.094	0.075	0.067	0.060	0.054	0.207	0.188	0.179	0.170	0.163
113	45103030	0.658	0.408	0.267	0.242	0.222	0.563	0.488	0.402	0.390	0.376
114	45201020	0.051	0.042	0.034	0.032	0.028	0.172	0.157	0.141	0.137	0.128
115	45202010	0.182	0.172	0.158	0.154	0.140	0.304	0.298	0.292	0.291	0.275
116	45202020	0.087	0.074	0.066	0.062	0.059	0.209	0.193	0.183	0.178	0.173
117	45202030	0.416	0.313	0.229	0.206	0.189	0.526	0.473	0.424	0.407	0.388
118	45203010	0.070	0.058	0.048	0.043	0.042	0.194	0.177	0.164	0.155	0.153
119	45203015	0.648	0.453	0.314	0.232	0.179	0.631	0.553	0.485	0.383	0.319
120	45203020	0.248	0.210	0.157	0.147	0.128	0.344	0.329	0.294	0.289	0.267
121	45203030	0.492	0.321	0.246	0.189	0.178	0.443	0.378	0.358	0.306	0.300
122	45204010	0.914	0.679	0.481	0.438	0.408	0.693	0.640	0.584	0.569	0.549
123	45301010	0.235	0.196	0.158	0.126	0.111	0.334	0.319	0.303	0.265	0.248
124	45301020	0.068	0.060	0.053	0.048	0.042	0.185	0.175	0.166	0.159	0.149
125	50101010	0.230	0.186	0.142	0.128	0.122	0.324	0.299	0.268	0.256	0.250
126	50101020	0.140	0.109	0.088	0.085	0.077	0.253	0.230	0.210	0.208	0.200
127	50102010	0.084	0.076	0.072	0.071	0.066	0.196	0.186	0.182	0.180	0.175

B.4 Root mean square deviations (RMSD)

Table B.5: RMSD

ID.	GICS code	AF	WWAF	MF	AWMF
1	10101010	0.014	0.033	0.118	0.207
2	10101020	0.278	0.340	0.410	0.495
3	10102010	0.290	0.307	0.269	0.283
4	10102020	0.192	0.262	0.105	0.123
5	10102030	0.483	0.584	0.624	0.736
6	10102040	0.184	0.227	0.500	0.571
7	10102050	0.629	0.745	0.773	0.856
8	15101010	0.188	0.498	0.154	0.182
9	15101020	0.149	0.290	0.210	0.310
10	15101030	0.420	0.726	0.513	0.787
11	15101040	0.177	0.318	0.230	0.427
12	15101050	0.612	0.662	0.748	0.789
13	15102010	0.059	0.133	0.048	0.063
14	15103010	0.253	0.253	0.246	0.183
15	15103020	0.687	1.063	0.399	0.561
16	15104010	0.103	0.198	0.097	0.217
17	15104020	0.049	0.043	0.100	0.113
18	15104030	0.307	0.352	0.339	0.404
19	15104040	0.494	0.688	0.522	0.753
20	15104045	0.332	0.540	0.369	0.642
21	15104050	0.072	0.069	0.132	0.108
22	15105010	0.247	0.555	0.273	0.503
23	15105020	0.063	0.105	0.059	0.074
24	20101010	0.020	0.024	0.122	0.148
25	20102010	0.726	0.705	0.800	0.714
26	20103010	0.012	0.165	0.071	0.069
27	20104010	0.201	0.210	0.276	0.340
28	20104020	0.018	0.065	0.126	0.088
29	20105010	0.244	0.419	0.296	0.516
30	20106010	0.130	0.069	0.216	0.203

Continued on next page

Table B.5 – *Continued from previous page*

ID.	GICS code	AF	WWAF	MF	AWMF
31	20106015	0.303	0.483	0.395	0.630
32	20106020	0.138	0.075	0.165	0.145
33	20107010	0.282	0.256	0.358	0.352
34	20201010	0.799	0.954	0.684	0.719
35	20201050	0.332	0.455	0.260	0.293
36	20201060	0.323	0.481	0.389	0.475
37	20201070	0.763	0.859	0.932	0.972
38	20201080	0.236	0.450	0.286	0.468
39	20202010	0.403	0.566	0.372	0.422
40	20202020	1.001	1.323	0.931	1.141
41	20301010	0.075	0.273	0.085	0.225
42	20302010	0.539	0.436	0.522	0.344
43	20303010	0.099	0.021	0.081	0.184
44	20304010	0.370	0.545	0.489	0.702
45	20304020	0.142	0.075	0.050	0.176
46	20305010	0.158	0.291	0.183	0.360
47	20305020	0.206	0.545	0.253	0.543
48	20305030	0.134	0.289	0.140	0.331
49	25101010	0.248	0.249	0.239	0.262
50	25101020	0.516	0.815	0.602	0.824
51	25102010	0.311	0.444	0.382	0.534
52	25102020	0.098	0.206	0.105	0.241
53	25201010	0.155	0.320	0.256	0.397
54	25201020	0.165	0.328	0.140	0.223
55	25201030	0.542	0.688	0.541	0.614
56	25201040	0.181	0.263	0.213	0.319
57	25201050	0.620	0.792	0.574	0.638
58	25202010	0.506	0.572	0.297	0.247
59	25202020	1.201	2.101	1.064	1.586
60	25203010	0.344	0.452	0.303	0.322
61	25203020	0.284	0.343	0.545	0.688
62	25203030	1.581	1.943	1.417	1.635
63	25301010	0.111	0.081	0.145	0.163

Continued on next page

Table B.5 – *Continued from previous page*

ID.	GICS code	AF	WWAF	MF	AWMF
64	25301020	1.003	1.163	0.694	0.696
65	25301030	0.607	0.752	0.798	0.895
66	25301040	0.365	0.750	0.340	0.640
67	25302010	0.003	0.227	0.344	0.235
68	25302020	0.020	0.125	0.256	0.296
69	25401010	0.605	0.694	0.745	0.810
70	25401020	0.685	0.848	0.598	0.658
71	25401025	0.378	0.495	0.172	0.199
72	25401030	0.509	0.695	0.470	0.567
73	25401040	0.018	0.008	0.101	0.150
74	25501010	1.424	1.630	1.321	1.411
75	25502010	0.223	0.229	0.326	0.293
76	25502020	0.188	0.294	0.259	0.445
77	25503010	0.822	1.012	0.699	0.742
78	25503020	0.097	0.189	0.035	0.063
79	25504010	0.074	0.111	0.155	0.246
80	25504020	0.060	0.117	0.055	0.132
81	25504030	0.433	0.630	0.524	0.686
82	25504040	1.030	1.158	0.737	0.712
83	25504050	0.219	0.175	0.275	0.227
84	25504060	0.067	0.073	0.205	0.294
85	30101010	0.308	0.660	0.407	0.733
86	30101020	0.078	0.159	0.235	0.438
87	30101030	0.020	0.185	0.097	0.020
88	30101040	0.590	0.751	0.769	0.900
89	30201010	0.286	0.465	0.363	0.599
90	30201020	0.009	0.009	0.094	0.086
91	30201030	0.067	0.128	0.083	0.130
92	30202010	0.235	0.272	0.530	0.654
93	30202030	0.075	0.252	0.114	0.052
94	30203010	0.117	0.116	0.117	0.139
95	30301010	0.167	0.221	0.151	0.222
96	30302010	0.100	0.279	0.050	0.151

Continued on next page

Table B.5 – *Continued from previous page*

ID.	GICS code	AF	WWAF	MF	AWMF
97	35101010	0.331	0.518	0.280	0.381
98	35101020	0.160	0.145	0.173	0.191
99	35102010	0.703	1.359	0.546	1.001
100	35102015	0.371	0.407	0.637	0.682
101	35102020	0.805	0.921	0.452	0.431
102	35102030	0.120	0.264	0.050	0.091
103	35103010	0.172	0.215	0.353	0.438
104	35201010	0.140	0.115	0.204	0.243
105	35202010	0.065	0.027	0.218	0.171
106	35203010	0.479	0.524	0.708	0.763
107	40402040	0.006	0.015	0.005	0.015
108	45101010	0.019	0.032	0.098	0.145
109	45102010	0.717	0.849	0.426	0.444
110	45102020	0.192	0.139	0.376	0.374
111	45103010	0.209	0.233	0.095	0.057
112	45103020	0.319	0.519	0.143	0.224
113	45103030	0.332	0.457	0.384	0.540
114	45201020	0.047	0.072	0.318	0.396
115	45202010	0.960	1.169	0.721	0.745
116	45202020	1.055	1.560	0.670	0.936
117	45202030	0.726	0.850	0.872	0.953
118	45203010	0.013	0.064	0.126	0.115
119	45203015	0.191	0.223	0.569	0.712
120	45203020	0.363	0.428	0.472	0.566
121	45203030	0.200	0.111	0.417	0.378
122	45204010	0.092	0.170	0.096	0.139
123	45301010	0.446	0.487	0.573	0.630
124	45301020	0.158	0.124	0.266	0.289
125	50101010	0.314	0.370	0.629	0.683
126	50101020	0.034	0.106	0.077	0.074
127	50102010	0.912	1.354	0.805	1.093

Chapter 4

Modelling country and group levels corporate default dependence: Evidence from the Euro area

4.1 Introduction

Over the last decades, there has been a steady increase in economic integration activities across the globe, and their importance has been widely recognized among policy-makers and researchers (see Flam, 1992; Summers, 1999; Crawford et al., 2005; Fiorentino et al., 2007; El-AGraa, 2011; among others), given the potential economic benefits that can be derived from trade, capital inflows, multinational cooperation and policy spillovers (Summers, 1999). However, it has been shown that economic integration may also have a negative impact on member states (see Tinbergen, 1954; Summers, 1999; Rose, 1999; Minford et al. 2015; and others). In general, the positive or negative impact may depend on the economic credentials of the members (see Aitken 1973; Baldwin, 1994; Rose, 1999; Cappelen et al., 2003; Baldwin and Krugman, 2004; Grimwade, 2007; Farole et al. 2011, among others).

The Eurozone offers some advantages to its members, such as (i) free movement of trade and capital; (ii) reduction transaction costs; (iii) elimination of exchange rate uncertainty; (iv) enhancement of price transparency; and (v) economic integration (see De Grauwe, 2010; Bak and Maciejewski, 2015). However, it no longer allows the member states to respond to asymmetric shocks independently, because of the loss in monetary sovereignty. The global financial crisis and the banking, sovereign debt and growth crises have seriously impacted on the economic integration process in the Euro area. Not only the PIIGS countries (Portugal, Ireland, Italy, Greece and Spain), but also Belgium and France were hit by the crises (Metiu, 2012; Arghyrou and Kontonikas, 2012; Ludwig, 2014) although to a lesser extent.[1] As a result, banks were more conservative with their lending activities, and a large reduction in loan supply was observed, with an impact on investment activities, job creation, and sale growth (Acharya et al., 2015). Since then business entities within the Euro bloc have struggled to survive, and the hazard rates of these businesses have been severely affected, due to their exposure to risk factors at country and group level.

In this chapter, we estimate the failure dependence of 1,422 public listed firms in 11 Eurozone countries over the period 1994Q1-2014Q4. The analysis is conducted at country level and group level. In particular, we consider the PIIGS and non PIIGS countries. Our choice of PIIGS countries follows the literature on the Euro crisis and also rely on a strong linkage between economic conditions and firm performance (see Bhattacharjee et al., 2009; Bonfim, 2009; Chen, 2010; Tang and Yan, 2010; Jacobson et al., 2013, among others). The non PIIGS countries are considered for comparison given the difference in firm and market characteristics. In addition, we consider three extra groups that are formed by including: only Belgium in the PIIGS (PIIGSB); only France in the PIIGS (PIIGSF); both Belgium and France (PIIGSBF) (see also Giordano et al., 2013). This allows us to establish to what extent the crises in the Euro area really affected Belgian and French firms performance along that of those firms in the PIIGS. The choice of considering these extra groups is mainly due to the impact that the Euro crisis had on Belgium and France (see previous paragraph).

We use a nested frailty model that accounts for two hierarchical clustering (see e.g. Sastry, 1997; Duchateau and Janssen, 2008; Wienke, 2011) within a multivariate framework of mixed effects Cox model (see Ripatti and Palmgren, 2000; Therneau and Grambsch, 2000; Martinussen et al., 2002; Therneau et al., 2003, among others). A cluster level-specific random effect, which is assumed to follow a Gaussian distribution, is considered for country and group level clustering, and firms in each country are exposed to country (internal) and group level (external) risk (unobserved) factors. Further, we also consider non-frailty and non-nested frailty models. The former, which do not account for unobserved factors, are used as benchmarks, while the latter, which account only for country level (internal) unobserved factors (firms are not exposed to any potential external risk factors), are used for comparison with the nested frailty models. As for the specification of the models, we select covariates from Shumway (2001), Duffie et al. (2007),

[1] It has been argued that a contagion effect has propagated from Belgium to France, as a result of the distressed bank Dexia.

and Bharath and Shumway (2008): distance to default probability, one-year trailing stock return, one year trailing market return, firm age, and 3 month T-bill rate.

The empirical analysis offers three main results. First, the distance to default probability covariate is positive and statistically significant, while the one trailing year stock return, one year trailing market return, and ln(age) variables are negatively statistically significant in all the models. In addition, the 3 month T-bill rate is insignificant in the case of non-frailty and non-nested frailty models. These results imply that: (i) an increase in distance to default probability pushes the firms towards a potential failure; (ii) older firms with high stock returns are less likely to experience failure, as compared to younger firms with lower stock returns; and (iii) the default rates decrease in the one year trailing market return, which seems to suggest that previous year's market performance tends to enhance firms' performance in the following years. Second, the failure dependence induced by firms' exposure to both country and group level unobserved factors is significantly larger than that observed with the non-nested frailty models. This is due to the fact that failure dependence due to firms' exposure to internal and external risk factors are appropriately captured with the nested frailty models. Third, models that account for the distance to default probability covariate tend to outperform their counterparts, since this covariate has a higher explanatory power in default rate models.

This study offers two contributions to the empirical literature. First, while previous papers focus on default clustering at either economy-wide or industry level within a country, we take a further step by looking at default clustering at country level and group level. In doing so, we are able to appropriately capture firms exposure to both internal and external risk factors, which may play an important role in corporate financial decision-making. Second, this study examines default clustering of public listed firms on 11 stock markets in the Euro area. To the best of our knowledge, this is the first study to examine default clustering at country and group levels within the Euro area.

The rest of the chapter is organised as follows. Section 4.2 reviews previous studies on the dynamics of corporate default risks in the Euro countries. Section 4.3 presents methodology and data. Empirical findings are presented in Section 4.4. Section 4.5 concludes the chapter.

4.2 Literature review

In this section, we review previous studies on corporate default that consider Euro countries.

Altman et al. (1994) compared classical statistical techniques such as linear discriminant analysis or logistic regression-based analysis with neural networks for failure classification and prediction. They employed 10 financial ratios of over 1000 Italian industrial firms categorised into healthy, vulnerable and unsound (distressed) groups, over the period 1982-1992. The results showed that both techniques produced over a 90% accuracy rate in terms of the classification of firms into groups and out-sample prediction. The authors suggested that models that combine both methods are likely to produce better predictions than those that employ the individual methods.

Using a different data set of Italian firms, Ciampi (2015) examined the nexus between corporate governance and default variation among 934 small enterprises (SEs) over the period 2008-2011. The authors found that CEO duality, owner concentration, and at most a 50% reduction in the number of outsiders on the board of directors have a negative impact on the failure rate of SEs. Besides, the author compared the predictions of the default accuracy using a model with financial ratio and corporate governance covariates and a model that uses only financial ratio covariates. The empirical results showed that the model which includes corporate governance factors tend to have comparatively higher accuracy rate.

Wallrafen et al. (1996) explored the performance of bankruptcy estimation and prediction models based on only neural networks (NN) and both neural networks and sequential genetic algorithms (SGA) on the basis of beta-error-"misclassification of solvent companies". The authors employed 73 financial ratios extracted from the financial statements of 6667 German corporations and found that, when NN and SGA are combined, the beta-errors of the estimation are reduced, as compared to those of the analyses that use only NN technique. Therefore, the paper concluded that models based on both techniques tend to produce lower misclassification errors as compared to those that use only neural networks-based methods.

In a related study, Rudorfer (1995) employed artificial neutral networks (ANN) to detect company failure using five balance sheet ratios of 59 healthy and 23 insolvent Austrian companies. Their approach revealed that a company with a high positive value of liabilities to total assets or quick assets to assets ratio tends to have a higher hazard of default, while the financially healthy companies are characterised by low liabilities to total assets, and quick assets to total assets ratios. The author further showed that accuracy of the estimates by ANN technique are pretty similar to those by the discriminant approach; and as such, concluded that the ANN approach is a preferred alternative for designing early warning systems for companies.

Zopounidis and Doumpos (1999) examined the potential of a new non-parametric approach called “multicriteria decision aid discrimination” (MCDA, henceforth) for the predictions of the failure rates among 80 Greek industrial firms. The performance of this approach is compared to that of discriminant and logistic analysis. The authors showed that the MCDA technique performs much better than the discriminant analysis, while it produces similar results to those by the logistic regression. The authors argued that their approach could be very helpful in making classification decisions (e.g., estimation of corporate default risk, credit administration problems, and portfolio selection).

Kaski et al. (2001) employed a self-organising map (SOM, hereafter) approach to analyse firm behaviour of 1342 active and 158 failed Finnish small medium enterprises(SMEs, hereafter). The authors used Fisher information matrix to calculate an SOM metric, which can be used to visually examine the bankruptcy of the SMEs. With the metric, they are able to visualize the current and future direction of the financial status of an SME; identify a wide range of the firms' behaviour. Another study on Finnish firms is by Laitinen (2007). The author employed a customised Linear discriminant analysis (LDA, hereafter), with the discriminant score function assumed to be uniformly rather than normally distributed, to examine correlation and failure classification for a set of 2092 failed and 63,072 active Finnish firms. The author found that the discriminant score follows neither a normal distribution nor a uniform distribution for a set of firm covariates. However, it is shown that the uniform distribution seems to approximate more accurately the correlations and produce higher failure classification accuracy rate than those of the normal distribution.

De Andrés et al. (2005) conducted a comparative analysis to examine the prediction performance of parametric (LDA or logit) and non-parametric (neural networks and fuzzy rule-based systems) techniques in discriminating between healthy and unhealthy commercial and industrial Spanish companies. The authors also employed a Monte Carlo simulation method to investigate sample size effects on the predictive accuracy of both the parametric and non-parametric methods. The empirical results showed that the neural networks and fuzzy rule-based system methods tends to produce more accurate estimates as compared to those by the LDA and logit methods.

Dewaelheyns and Van Hulle (2006) used subsidiary level and group level data set of large non-financial Belgian limited liability firms over the period 1996-2001. The authors argued that business groups in Europe are central to the respective subsidiaries performances, and hence models which treat firms as standalone entities are likely to produce less accurate results. Drawing on default literature, their model exhibited a higher level of prediction and classification accuracy. The authors also examined the survival probability of the subsidiaries when they receive any support from the business group. Their results show that the survival probability tend to increase when the subsidiaries are into the core business activities of the business group.

du Jardin and Séverin (2012) compared the failure prediction performance of conventional approaches such as discriminant analysis, logistic regression, Cox models and neural networks with those of Kohonen map using a set of French companies with at least 6 years of operation. In particular, the authors examined the estimation and forecasting stability of the two techniques by extracting the financial performance of the firms over a given period. du Jardin and Séverin (2012) showed that the Kohonen map-based default models tend to be relatively more stable over time than those of the conventional tools, and these models may help financial institutions to minimize the margin of error in their risk management decisions. A different model to study the default rate of French firms over the period 2003-2012 is used by du Jardin (2015). The author developed a default prediction model that explicitly control for how a firm moves towards failure few distance away from the actual default. For the empirical analysis, the authors also considered discriminant analysis, logistic regression, neural networks, survival analysis, and SOM (the benchmarks). The results showed that the model developed by du Jardin (2015) tends to outperforms the benchmarks for the predictions of the default rates up to a 3 year forecasting horizon.

Using 10599 non-failed and 1582 failed contractors in the Portuguese construction industry over the period 2008-2010, Horta and Camanho (2013) proposed a novel model for predicting company default. The novelty of their model is the use of financial and strategic variables that accurately capture the important specifics of the construction firms, the use of support vector machine (SVM) approach, and the improvement of the estimates from SVM by using random oversampling and random undersampling methods. The authors showed that their approach is very robust for predicting failure within the construction industry. More specifically, based on the receiver operating characteristic (ROC) curve measures, the estimates of the SVM are more accurate as compared to those of the logistic regression. The authors concluded that the SVM approach produces more accurate default forecasts within the construction industry.

These studies improved on the classical methods of firm default prediction (discriminant, logit, and probit analyses) in many respects. For instance, Altman et al. (1994) and Rudorfer (1995) considered a possible non-linear relationship between default and firms characteristics by using neural networks. However, Wallrafen et al. (1996) augmented neural networks with sequential genetic algorithm showed an improved classification of default and non-default firms. De Andrés et al. (2005) further argued that combining neural network-based approach with fuzzy rule-based technique tends to produce more accurate estimates as compared to the classical methods, while Horta and Camanho (2013) argued that support vector machine approach produces more accurate results as compared to those of logistic regression. Using a different approach, Kaski et al. (2001), du Jardin and Séverin (2012) and du Jardin (2015) used graphical methods (self-organising maps and Kohonen maps) to determine the future direction of financial status of firms over a time horizon and their empirical analysis revealed that these methods tend to produce stable estimates of the parameters over time.

While all the above works explicitly improve on the traditional techniques, none of them explicitly adjusts for the effects of unobserved factors on active firms, especially distressed market periods. Furthermore, they do not explicitly assume that firms are exposed to external risk factors, as countries are treated as standalone economic entities in the empirical analyses. In addition, most of the techniques and results are purely data-driven with no relevant theoretical basis. As such, this study fills the gap, and investigates the effects of country-based and group level risk factors on firms within the Eurozone.

4.3 Methodology and data

In this section, we proceed as follows. First, we briefly present the non-nested frailty model and the non-frailty model. Second, we present the two level nested frailty model: one for a country level and the other for a group level. Third, we describe the data used in the empirical analysis.

4.3.1 Non-nested frailty model

Let T_{ij} and δ_{ij} respectively be the event time and event indicator (censoring indicator) of firm i listed in country j among q countries. The indicator δ_{ij} takes the value 1 if T_{ij} is a failure time and 0 otherwise. Suppose that the data set of firm i follows a shared frailty model in the context of an extended Cox Proportional Hazard (PH) model. The hazard rate of the firm is defined as follows (see Hougaard, 2000; Ripatti and Palmgren, 2000 Therneau et al., 2003; Duchateau and Janssen, 2008; amongst others):

$$\lambda_{ij}(t) = \lambda_0(t) u_i exp(X_{ij}(t)\beta), \tag{4.1}$$

where $\lambda_{ij}(t)$ is the hazard rate of firm i listed in country j with a vector of covariates $X_{ij}(t)$, which have a vector of parameter estimates β. The frailty term u_i, which is shared amongst firms listed in country j, acts multiplicatively on the hazard rate. The baseline hazard function $\lambda_0(t)$ is assumed to be unknown, which makes the hazard rate in equation

(4.1) semi-parametric.

The non-frailty model is derived from equation (1) by setting the frailty term $u_i = 1$:[2]

$$\lambda_{ij}(t) = \lambda_0(t)exp(X_{ij}(t)\beta). \tag{4.2}$$

Following the literature on the penalised partial likelihood (PPL, hereafter) (see e.g. Ripatti and Palmgren, 2000; Therneau and Grambsch, 2000; Duchateau et al., 2002; Therneau et al., 2003), we present an alternative formulation of equation (4.1):

$$\lambda_{ij}(t) = \lambda_0(t)exp(X_i\beta + Z_iw), \tag{4.3}$$

where $w_i = log(u_i)$ and Z is a matrix of q indicator variables, with $Z_{ij} = 1$ if firm i is listed in country j, and 0 otherwise. In model (4.3) each firm is listed in only one country. In other terms, cross listing of firms is not allowed.

In order to incorporate time varying covariates in the estimation of equation (4.1), we employ the counting process input style of Andersen and Gill (1982). As a result, the pair (T_{ij}, δ_{ij}) for firm i listed in country j is substituted by $(N_i(t), Y_i(t))$, where $Y_i(t)$ assumes value 1 if firm i is still active and 0 otherwise, and $N_i(t)$ is the number of events in the period $(t_l, t_{t+1}]$ for firm i, with t_l and t_{l+1} being the beginning and the ending time of the interval. We follow McGilchrist and Aisbett (1991) and McGilchrist(1993) in such a way that the random effect w in equation (4.3) is normally distributed on the log-scale, and the parameters β and w are estimated by maximizing the PPL:

$$PPL = PL(\beta, w; data) - g(w; \theta), \tag{4.4}$$

where PL is defined as the log of the classical Cox partial likelihood conditioned on the data set:

$$PL(\beta, w) = \sum_{i=1}^{n} \int_0^{\infty} \Big[Y_i(t)(exp(X_i\beta + Z_iw) - log\Big(\sum_k Y_k(t)(exp(X_k\beta + Z_kw)\Big)\Big] dN_i(t), \tag{4.5}$$

and the penalty term is defined by

$$g(w; \theta) = \frac{1}{2\theta} \sum_{j=1}^{q} w_i^2, \tag{4.6}$$

where θ is the variance of the log-frailty or random effect.[3] For a given value of the

[2]For the maximum likelihood derivation of equation (4.2), see Appendix A

[3]For details on penalised partial likelihood of a shared frailty model, see Ripatti and Palmgren (2000)

variance estimate θ, we use the expansion and approximation of Ripatti and Palmgren (2000) to derive a modified likelihood defined as:

$$
\begin{aligned}
l_m(\beta,\theta) =& -\frac{1}{2}\log(|D|) + \log\Big(\int exp\Big[PL(\beta,w) - \frac{1}{2}w'D^{-1/2}w\Big]dw\Big) \\
&\approx PL(\beta,\tilde{w}) - \frac{1}{2}\log\Big(\tilde{w}'D^{-1/2}\tilde{w} + log|D|) + \log(|H_{22}(\beta,\tilde{w})\Big), \qquad (4.7)
\end{aligned}
$$

where $D = \theta I$ is a diagonal matrix and I is an identity matrix of order $q \times q$; q is the number of countries in the sample, and $g(w;\theta) = \tilde{w}'D^{-1}(\theta)\tilde{w}$. The term $\tilde{w} = \tilde{w}(\beta,\theta)$ solves the following equation

$$
\sum_{i=1}^{n}\int_0^{\infty}(Z_{ij} - Z_j(t))dN_i(t) - D^{-1}(\theta)\tilde{w} = 0 \qquad (4.8)
$$

We maximize the likelihood in equation (4.7) over the parameters using the "$coxph$" procedure in the "survival" package in R (see Therneau, 2015).

4.3.2 Nested frailty model

Our sample comprises of s clusters (groups), and in each group there are n_i subclusters (countries). Further, each country contains n_{ij} members (firms) (see Duchateau and Janssen, 2008). In this setting, firms are located within countries, and countries are nested in groups. The nested frailty model is given by:

$$
\begin{aligned}
\lambda_{ijk}(t) =& \lambda_0(t)u_i z_{ij} exp(X_{ijk}(t)\beta) \\
=& \lambda_0(t) exp(X_{ijk}(t)\beta + w_i + v_{ij}), \qquad (4.9)
\end{aligned}
$$

where $\lambda_{ijk}(t)$ is the hazard rate at time t of firm $k = 1, ..., n_{ij}$ in country $j = 1, ..., s_i$ located in group $i = 1, ..., s$. The term $\lambda_0(t)$ is the baseline hazard function at time t and β is a $p-$dimensional parameters of the set of covariates, $X_{ijk}(t)$. In addition, $w_i = logu_i$ is the random effects term for group i, whilst $v_{ij} = logz_{ij}$ is the random effects term of country j nested in group i. We define T_{ijk} as the event time of firm k listed in country j located in group i with a corresponding censoring indicator, δ_{ijk}. The latter takes value 1 if T_{ijk} is a failure time and 0 otherwise. The nested model is developed within a multivariate shared frailty framework (see Section 4.3.1) and equation (4.9) can be re-written as follows (see

and Therneau et al. (2003).

Ripatti and Palmgren, 2000; Martinussen et al., 2002; Therneau et al., 2003; among others):

$$\lambda(t) = \lambda_0(t) exp(X\beta + Zb), \qquad (4.10)$$
$$b \sim G\Big(\theta, \sum(\theta)\Big),$$

where X and Z are the time-varying covariate matrices for the fixed and random effects, respectively, and β and b are the vectors of fixed and random effects coefficients with dimensions q. The non-negative term λ_0 is the baseline hazard function and it is assumed to be unknown. The coefficients of equation (4.10) are still possible to be estimated without knowing the shape of λ_0. The random effects distribution G is a multivariate Gaussian distribution with zero mean and variance matrix $\sum$, which is a function of a vector of the parameters θ. Following Therneau and Grambsch (2000) and Therneau et al. (2003), we define the log penalised partial likelihood function as follows:

$$PPL(\beta, b, \theta) = l(\beta, b) - g(b, \theta), \qquad (4.11)$$

where the penalty function $g(b, \theta) = b' \sum^{-1}(\theta) b/2$. The term $l(\beta, b)$ is called the partial likelihood (PL) in a Cox setting (see Therneau, 2015) for any given value of β and b, and is defined as:

$$l(\beta, b) = \sum_{k=1}^{n_{ij}} \int_0^{\infty} \Big[Y_k(t)\eta_k(t) - log\Big(\sum_j Y_j(t)\eta_j(t)\Big)\Big] dN_k(t), \qquad (4.12)$$

where $\eta_k(t) = X_k(t)\beta + Z_k(t)b$ is the linear score for firm k at time t, $X_k(t)$ and $Z_k(t)$ are the k^{th} rows of the covariate matrices X and Z, respectively. In other words, the above row matrices are the data set for firm k in country j. The term $Y_k(t)$ describes the surviving firms (or firms still at risk of default) which takes value 1 when firm k is active at time t and 0 otherwise. Equation (4.11) can then be re-written as :[4]

$$PPL(\beta, b, \theta) = \sum_{k=1}^{n_{ij}} \int_0^{\infty} \Big[Y_k(t)\eta_k(t) - log\Big(\sum_j Y_j(t)\eta_j(t)\Big)\Big] dN_k(t) - \frac{b' \sum^{-1}(\theta) b}{2}. \qquad (4.13)$$

The estimates of β and b, $\hat{\beta}$ and $\hat{b}$, are obtained by solving the following the score equations (see Therneau et al., 2003):

[4] For detailed treatment, refer to Therneau and Grambsch (2000) and Therneau et al. (2003).

$$\frac{\partial PPL}{\partial b_j} = \sum_{i=1}^{n} \int_0^{\infty} (Z_{ij} - Z_j(t)) dN_i(t) - \frac{\partial g(b;\theta)}{\partial b_j}, \tag{4.14}$$

$$Z_j(t) = Z_j(\beta, b, t) = \frac{\sum Z_k j Y_k [X_k \beta + Z_k b]}{\sum Y_k [X_k \beta + Z_k b]}. \tag{4.15}$$

We also obtain the integrated partial likelihood (IPL) by integrating out the random effects as obtained below (Therneau, 2015):

$$IPL = \frac{1}{(2\pi)^{q/2} |\sum(\theta)^{1/2}|} \int PPL(\beta, \theta) exp\Big(- b^{'} \big(\sum \big)^{-1} (\theta) b/2 \Big) db, \tag{4.16}$$

where q is the number of random effects. We estimate the parameters using the "$coxme$" package in R by Therneau (2015).

4.3.3 Data

Our data are drawn from DataStream and Worldscope for public listed firms in 11 member states of the Eurozone for the period 1994Q1-2014Q4. The sample is comprised of 1,422 firms: 905 active firms, 398 failed firms and 119 acquired or merged firms, and this translates into 71,680 quarterly firm observations. The countries include Austria, Belgium, Finland, France, Germany, Greece, Ireland, Italy, Netherlands, Portugal, and Spain. The choice of these countries is based on data availability.[5] Table 4.1 presents firms' status at country level across the 11 selected members of the Eurozone.

The definition of firm failure may differ across all the member states of the Euro area. Therefore, for the sake of uniformity, we follow Altman and Narayanan (1997) who defines the following as failure: (i) filing by a company; (ii) bond default; (iii) bank loan default; (iv) delisting of a company; (v) government intervention via special financing; and (vi) liquidation.[6] We select failed, and acquired or merged firms from the DataStream "DEAD" category for each country in conjunction with other sources (e.g. Bloomberg bankruptcy segment). For instance, the DataStream items "DEADGR", "DEADBD" and "DEADFR" are the categories for dead firms in Greece, Germany and France, respectively.

[5]These countries may have some accounting information disclosure differences, but Worldscope adjusts the variables for these differences.

[6]For delisting of a company, we cross check the reasons for delisting at other sources. These reasons include mergers, acquisitions and some of the reasons already stated.

Table 4.1: Active, failed and merged or acquired firms within the Eurozone

ID	Country	Active firms	Failed firms	Merged/Acquired firms	Total
1	Austria	33	41	4	78
2	Belgium	57	29	10	96
3	Finland	39	22	14	75
4	France	200	51	21	272
5	Germany	199	44	10	253
6	Greece	38	47	18	103
7	Netherlands	78	56	10	144
8	Ireland	31	24	2	57
9	Italy	111	32	13	156
10	Portugal	37	27	4	68
11	Spain	82	25	13	120
	Total	905	398	119	1422

4.3.3.1 Dependent variable

In duration models, the dependent variable is the time taken for a subject to experience either a non-failure or failure event. Time to event is usually specified with the corresponding event indicator which takes value 1 for failure event and 0 otherwise. To incorporate time varying covariates, we use the counting process input style following Andersen and Gill (1982). For instance, suppose it takes 6 years for a firm to experience an event. For a failure event, we construct the intervals $(0,1]$, $(1,2]$, $(2,3]$, $(3,4]$, $(4,5]$, and $(5,6]$ for year 1, 2, 3, 4, 5, and 6, respectively. The event indicator is 0 for the years 1, 2, 3, 4, and 5 when the firms is still active, but takes 1 for the 6th year, when the firm failed. We can therefore simply reconstruct the intervals as follows: $(0,1;0]$, $(1,2;0]$, $(2,3;0]$, $(3,4;0]$, $(4,5;0]$, and $(5,6;1]$, where the first and second values are the beginning and the end of the year, and the last one is the event indicator. For example, $(2,3;0]$ indicates the dependent variable for the third year, where 2 and 3 are the beginning and end of the third year, and the third value 0 is the event indicator, since the firm is still traded at the end of the third year. For non-failure event, e.g. where the firm is delisted as a result of merger and acquisition activities, we have $(0,1;0]$, $(1,2;0]$, $(2,3;0]$, $(3,4;0]$, $(4,5;0]$, and $(5,6;0]$. The event indicator is 0 for all the intervals since the firm is censored as a result of a non-failure event.[7]

[7] For details on counting process for Cox regression, refer to Andersen and Gill (1982).

4.3.3.2 Independent variables

We employ some widely used covariates in the empirical literature of corporate failure, given their explanatory power (see Shumway, 2001; Duffie et al., 2007; Duffie et al., 2009; Duan et al., 2012; Lando et al., 2013; Qi et al., 2014; Azizpour et al., 2015, among others). First, we use the 3 month T-bill rate, which is a measure of short-term interest rates. Second, we consider the one year trailing stock return, which is good predictor of firm failure (see Shumway, 2001), and is constructed by cumulating monthly stock returns. Third, we use the one year trailing market return, which is a measure the overall market performance, and is constructed by cumulating monthly market returns. Fourth, the distance to default probability is used as a probabilistic measure of volatility adjusted leverage. In constructing this measure, we follow Bharath and Shumway (2008): firms with higher probabilities are close to default, whilst firms with lower probabilities are far from default. Lastly, we consider the age of a firm to test whether older firms are less likely to fail than the younger ones (see Gong et al., 2004; George, 2005; Aldrich and Ruef, 2006; Wiklund et al., 2010).[8]

Table 4.2 presents the descriptive statistics for the covariates used for estimating the parameters of the non-frailty, non-nested frailty, and nested frailty models.

Table 4.2: Descriptive statistics

Variable	Mean	Std. Dev	Min.	25th P.	Median	75th P.	Max.
Distance to default prob.	0.122	0.176	0.000	0.002	0.046	0.179	1.000
Stock return (%)	12.195	46.398	-76.015	-15.195	4.340	32.028	200.690
Market return (%)	12.263	29.447	-66.066	-9.064	16.115	29.722	178.882
ln(age)	1.875	0.816	0.000	1.386	2.079	2.565	2.996
3 month T-bill rate (%)	4.418	2.528	1.123	2.598	3.788	5.211	12.144

Notes: The terms 25th P. and 75th P. represent 25th and 75th percentiles, respectively.

The distance to default probability has the minimum and maximum values of 0.000 and 1.000, respectively, with a mean of value 0.122 and a standard deviation of value 0.176. The stock return falls within the range -76.015% and 200.690%, while the minimum and maximum values of the market return variable are -66.066% and 178.882%, respectively. The stock and market returns have approximately a mean value of 12%, but the former varies more about the mean than the latter. Additionally, the natural of firm age is bounded by (0.000, 2.996) since the firm age falls within the interval [1, 21]. The 3 month T-bill rate ranges from 1.123% to 12.144%.

[8] For the definition of firm age, refer to Chapter 2.

4.4 Empirical analysis

This section presents the empirical results using using the non-frailty, non-nested frailty, and nested frailty models. First, we estimate the parameters of the non-frailty and the non-nested frailty models, respectively. We compare the estimates of the non-frailty model with those of the non-nested frailty model with the aim of showing the importance of accounting for frailty factors at the country level, especially during distressed market periods. Second, we estimate the nested frailty model and compare its performance with that of the non-nested model to show the importance of frailty factors, not only at country level, but also at the group (Euro) level. Third, we compute total (country level plus group level) riskiness of firms in order to measure how firms are affected by country and group level unobserved factors.

In our analysis (see also Introduction), we consider the PIIGS countries against the non PIIGS along with three extra groups that are formed by including: only Belgium in the PIIGS (PIIGSB); only France in the PIIGS (PIIGSF); and both Belgium and France (PIIGSBF). In other words, we extract the country and group level frailty factors that affect listed firms for following pair of groups: (i) PIIGS versus non-PIIGS, (ii) PIIGSB versus non-PIIGSB, (iii) PIIGSF versus non-PIIGSF, and (iv) PIIGSBF versus non-PIIGSBF.

4.4.1 Non-frailty and non-nested frailty models

In the regression analysis, we employ distance to default probability, one year trailing stock return, one year trailing market return, ln(age), and 3 month T-bill rate as covariates. The diverse specifications of the non-frailty model, that do not control for potential default correlations induced by unobserved factors, are reported in Table 4.3. For model 1, the hazard rate is assumed to be a function of only distance to default probability. The results in Table 4.3 confirm that this covariate is a good predictor (see Bharath and Shumway, 2008). For models 2 and 4, we test whether other covariates may contribute to predicting hazard rates.

Model 2 uses the distance to default probability along with stock return, market return, and log of firm age. Model 4 adds the 3 month T-bill rate to the specification of model 2. All the covariates are significant with the expected sign in both models, with the exception of 3 month T-bill rate. The significance of these covariates show that the distance to default probability is a very important predictor, but not sufficient for the prediction of failure rates (see also Bharath and Shumway, 2008). This implies that, for an appropriate default model specification, the distance to default probability should be augmented with suitable

Table 4.3: Non-frailty specifications. Models without random effects

	Dependent variable: Time to event			
	Model 1	Model 2	Model 3	Model 4
Distance to default prob.	1.315***	0.998***		0.997***
	(0.229)	(0.252)		(0.253)
Stock return		−0.326**	−0.556***	−0.336**
		(0.145)	(0.142)	(0.145)
Market return		−0.767**	−0.879***	−0.861***
		(0.313)	(0.318)	(0.317)
ln(age)		−0.377***	−0.361***	−0.372***
		(0.082)	(0.081)	(0.082)
3 month T-bill rate			6.616	6.288
			(4.379)	(4.366)
Log likelihood	−2,612.903	−2,595.912	−2,629.264	−2,594.950
LR test	27.915***	61.897***	50.056***	63.821***
	[0.000]	[0.000]	[0.000]	[0.000]
Wald test	33.060***	66.450***	49.960***	68.770***
	[0.000]	[0.000]	[0.000]	[0.000]
Score (Logrank) test	33.573***	67.639***	50.397***	70.057***
Pseudo-deviance	5225.806	5191.824	5258.528	5189.900
AIC	5227.806	5199.824	5266.528	5199.900
AICC	5227.827	5200.039	5266.743	5200.224
BIC	5228.087	5200.948	5267.652	5201.305

Notes: The efron approximation is used to control for ties in the event times of firms. The standard errors and p-values are in round and square brackets, respectively. *** and ** denote significance at the 1% and 5% level, respectively. The Wald, LR (log-likelihood ratio), and Score are global test statistics. The tests compare a model with and without covariates. The pseudo-deviance are used to compare the overall model fit of nested models, while the Akaike information criterion (AIC), corrected Akaike information criterion (AICC), and Bayesian information criterion (BIC) measures are used to compare either nested or non-nested models.

covariates. Model 3, which we consider to perform a confirmatory test on the distance to default probability, also shows the insignificance of the 3 month T-bill rate, while the other covariates are significant.

In order to compare the models in Table 4.3 in terms of goodness of fit, we use both the pseudo-deviance and information criteria measures. For the former criterion, we have:

$$Pseudo\text{-}deviance = -2loglik_A + 2loglik_B, \tag{4.17}$$

where $-2loglik_A$ and $2loglik_B$ are the deviance statistics for generic models A and B, respectively. The statistic in (4.17) says how model A perform worse than the supposed best model B, and it can be used in case of nested models. A nested model, in this regard, is a model can be obtained from another model (full model) by imposing restriction(s) on the coefficient(s) of the full model. The statistic follows a chi-square distribution (χ^2), with the degrees of freedom being the difference between the number of parameters in model A and that in model B.

As for the information criteria, we consider the Akaike information criterion (AIC), the corrected Akaike information criterion (AICC), and the Bayesian information criterion (BIC).[9] As a rule of thumb, the lower the values of these information criteria, the better the fit. The information criteria are suitable for both non-nested and nested models.

Since model 2 nests model 1, and model 4 nests model 2, we use the pseudo-deviance to compare the performance of these models. In case of models 1 and 2 (the latter nests the former), the pseudo-deviance statistic is equal to $33.982(5225.806-5191.824)$, with 3 degrees of freedom. Since the value of the statistic is greater than the critical value, $\chi^2_{0.001(3)} = 16.266$, the null hypothesis that the coefficients of stock return, market return and ln(age) are all equal zero can be rejected. This implies an improvement of the fit due to the inclusion of these covariates, which makes model 2 the best candidate. Similarly, the values of this statistic for models 1 and 4 is 35.906, which is larger than the critical value $\chi^2_{0.001(4)} = 18.467$. Therefore, model 4 fits the data better than Model 1. Instead, a different result is found for models 2 and 4. Here, the values of the pseudo-deviance statistic, 1.924, is smaller than the critical value, $\chi^2_{0.001(1)} = 10.828$; there is no improvement in terms of fit if one adds 3 month T-bill rate to model 4. Likewise, for models 3 and 4, the test statistic is 68.628 with $\chi^2_{0.001(1)} = 10.828$. The significance of this test suggests that model 4 fits data better than model 3.

When considering non nested models (models 2 and 3), the results related to the

[9] For details on the information criteria, refer to Section 2.3 in Chapter 2.

information criteria AIC, $AICC$, and BIC, show that Model 2 is the best model when random effects are not taken into account.

Using model 2 (the best model), we further explore the effects of covariates on the hazard rates (expected time to default) using the transformation $100(e^{\beta} - 1)$, where β is the coefficient of a given covariate. The coefficient of the distance to default probability covariate is 0.998 and produces the value $100(e^{0.998} - 1) = 171.285$. This implies that a unit increase in the distance to default probability variable leads to 171.285% increase in the instantaneous rate of default, holding other factors constant. Likewise, an increase in the stock return and the market return decrease the hazard of failure by 27.819% and 53.560%, respectively. Hence, an increase in these covariates causes the expected time to failure to increase. For a 1-year increase in the age covariate, the rate of default decreases by 31.408%, leading to an increase in expected time to default. It should be noted that these results are susceptible to the effects of unobserved risk factors.

In Table 4.4, we report the estimates of the counterparts of the non-frailty models. Models 5-8 account for the influence of unobserved factors.[10] The distance to default probability covariate is significant in models 5, 6 and 8, and stock return, market return and ln(age) are all negative and statistically significant in models 6-8. The 3 month T-bill rate is negative and insignificant in models 7-8.

Evaluating models 5-8 using the information criteria, model 6 turns to be the best model. When examining the percentage effects of the covariates on the instantaneous rate of default (using the transformation $100(e^{\beta} - 1)$) in the presence of unobserved risk factors, the following emerge. An increase in the distance to default probability variable causes a decline in the expected time to failure. Similarly, a 1-unit rise in the stock return and market return covariates leads to 25.770% and 42.822% decline in the instantaneous rate of failure, respectively, while the rate decreases by 31.887% for 1 additional firm age. Therefore, the expected time to failure rises with an increase in stock return, market return, and firm age covariates. These trends are robust to the effects of only country-based unobserved risk factors.

The above results have some implications for the Eurozone countries. First, the distance to default probability shows high explanatory power in hazard rate models for the Euro countries. This reveals that firms that usually exhibits averagely higher distance to default probability are more prone to experience failure within the Eurozone. Second, a rise in stock return and market return increases the expected time to default. This out-

[10]For instance, model 5 is a non-nested frailty model that is derived from model 1 by controlling for the effects of unobserved factors. Likewise, Model 6 is obtained from Model 2, and so on. Models 5-6 are more market driven, while models 7 and 8 include a mix of market driven and macroeconomic covariates.

Table 4.4: Non nested frailty specifications. Models with random effects

	Dependent variable: Time to event			
	Model 5	Model 6	Model 7	Model 8
Distance to default prob.	1.221***	0.929***		0.925***
	(0.231)	(0.255)		(0.255)
Stock return		-0.298**	-0.490***	-0.291**
		(0.143)	(0.140)	(0.144)
Market return		-0.559**	-0.482*	-0.501*
		(0.276)	(0.290)	(0.290)
ln(age)		-0.384***	-0.375***	-0.386***
		(0.082)	(0.081)	(0.082)
3 month T-bill rate			-3.814	-3.125
			(4.960)	(4.968)
LogLik.(Fitted)	-2572.592	-2557.045	-2588.069	-2556.648
LogLik.(Integrated)	-2586.790	-2571.050	-2602.738	-2570.852
Integrated LR test	80.140***	111.620***	103.110***	112.020***
	[0.000]	[0.000]	[0.000]	[0.000]
Penalized LR test	108.540***	139.630***	132.450***	140.430***
	[0.000]	[0.000]	[0.000]	[0.000]
Pseudo-deviance	5173.580	5142.100	5205.476	5141.704
AIC	5175.580	5150.100	5213.476	5151.704
AICC	5175.601	5150.315	5213.691	5152.028
BIC	5175.861	5151.224	5214.600	5153.109
Dependence	0.185	0.179	0.203	0.186

Notes: LogLik.(Fitted) and LogLik.(Integrated) are the fitted and integrated likelihoods due to unobserved factors, respectively. The terms Integrated LR and Penalized LR denote the unobserved factors-adjusted integrated and penalized likelihood ratio tests, respectively. For other details, see Table 4.3.

come seems to suggest that firms listed within the Euro area with a consistent increase in their returns are less likely to move towards a failure point, and performing markets tend to enhance the survival rate of such firms, as compared to those of averagely decreasing stock returns. Third, the significance of age in our models reveals that older firms in Euro area are less likely to fail than the younger ones. This may be due to the liability of newness (see Stinchcombe, 1965; Baum, 1996; Aldrich and Ruef, 2006; Wiklund et al., 2010, among others), as older firms may have more business contacts, better understanding of the dynamics of the business environment and more robust organisational structure. Further, older firms generally satisfy various regulatory requirements (see Nelson and Winter, 1982; Baum, 1996; Gong et al., 2004 George, 2005, among others). Finally, the lack of significance of the 3 month T-bill rate in our models indicate that the monetary authorities do not play a significant role in influencing the hazard rate of firms in the Eurozone during the period under investigation.

When comparing models in Tables 4.3 and 4.4, the following emerges. While the explanatory power of the covariates seem to be similar across non frailty and non nested frailty specifications, those models that account for potential failure rates correlations (see Table 4.4) seem to fit the data better than the other ones (see Table 4.3). For this reason, we use model 6 (the best unobserved factor specifications) as the standard model for the analysis in Section 4.4.2, where nested frailty models are considered.

4.4.2 Nested frailty models

The results in Section 4.4.1 are obtained under the hypothesis that firms are only exposed to country level (internal) unobserved factors. This implies that the potential group level (external) risk factors induced by financial and debt crises are completely ignored. Further, since the crises hit the Euro countries in different ways, the external factors may play here a relevant role. As such, it is worth investigating how the external factors induced by the crises may have affected the default rates of firms in the four groups of countries previously mentioned. We do this by assuming that countries in a group share similar characteristics due to the prevailing macroeconomic and firm-specific factors. In particular, we assume similar trends in firms' distance to default probabilities and stock returns for each group of countries, and run two regressions for each of the four pair of countries (PIIGS vs. non PIIGS; PIIGSB vs. non PIIGSB; PIIGSF vs. non PIIGSF, and PIIGSBF vs. non PIIGSBF) (for example D_{PG} and S_{PG} denote the regressions for the PIIGS countries based on similar trends in firms' distance to default probability and stock return, respectively). We use model 6 (see Table 4.4) as our standard model.

Table 4.5: Nested frailty models: PIIGS and PIIGSBF groups

	Dependent variable: Time to event			
	PIIGS		PIIGSBF	
	D_{PG}	S_{PG}	D_{BF}	S_{BF}
Distance to default prob	1.338***	0.861***	1.238***	0.891***
	(0.278)	(0.259)	(0.289)	(0.259)
Stock return	-0.266*	-0.311**	-0.242	-0.316**
	(0.145)	(0.147)	(0.148)	(0.148)
Market return	-0.725**	-0.744**	-0.731**	-0.763**
	(0.319)	(0.311)	(0.338)	(0.311)
In(age)	-0.373***	-0.356***	-0.361***	-0.397***
	(0.084)	(0.096)	(0.089)	(0.096)
LogLik.(Fitted)	-1844.363	-2490.691	-1234.067	-2468.146
LogLik.(Integrated)	-2561.092	-2592.953	-2586.009	-2592.104
Integrated LR test	131.540***	67.82.060***	81.700***	69.510***
	[0.000]	[0.000]	[0.000]	[0.000]
Penalized LR test	1564.990***	272.340***	2785.590***	317.430***
	[0.000]	[0.000]	[0.000]	[0.000]
Psuedo-deviance	5122.184	5185.906	5172.018	5184.208
AIC	5130.184	5193.906	5180.018	5192.208
AICC	5130.399	5194.121	5180.233	5192.423
BIC	5131.308	5195.030	5181.142	5193.332
Dependence	1.919	0.283	4.325	0.348

Notes: The standard errors and p-values are in round and square brackets, respectively. D_{PG} and S_{PG} indicate the models for PIIGS with similar trends in terms of the distance to default and stock returns respectively, whereas D_{BF} and S_{BF} are the models for the PIIGSBF.

The empirical results are illustrated in Tables 4.5 and 4.6. In all the models, almost all the regressors are significant with the expected signs. For example, in the model D_{PG}, the coefficient of distance to default probability is positive and significant, and those of the stock return, market return, and In(age) are negative and significant. While these results do not differ from those in Table 4.4, measures of dependence have improved considerably, regardless of the specification of the models.

In Table 4.7, we report results related to the risk scores and the level of riskiness of firms within countries, and PIIGS and non-PIIGS group of countries. In the event of failure clustering, the country (group) score shows how firms are likely to fail either faster or slower. As such, we use value 1 (expected value of frailty) as a threshold value for gauging riskiness. A risk-score large than 1 implies more riskiness, while a score lower than 1 is considered less riskiness. Examples of more risky countries are Austria, Finland, Greece, Ireland, Netherlands and Portugal, while the less risky countries are Belgium,

Table 4.6: Nested frailty models: PIIGSB and PIIGSF groups

	Dependent variable: Time to event			
	PIIGSB		PIIGSF	
	D_B	S_B	D_F	S_F
Distance to default prob	1.331***	0.847***	1.349***	0.876***
	(0.277)	(0.258)	(0.277)	(0.258)
Stock return	-0.272*	-0.320**	-0.262***	-0.317**
	(0.145)	(0.147)	(0.145)	(0.149)
Market return	-0.712**	-0.732**	-0.757***	-0.752**
	(0.318)	(0.311)	(0.321)	(0.311)
ln(age)	-0.369***	-0.356**	-0.378***	-0.391***
	(0.084)	(0.096)	(0.084)	(0.097)
LogLik.(Fitted)	-1839.700	-2487.728	-1788.692	-2451.328
LogLik.(Integrated)	-2560.389	-2592.065	-2563.549	-2592.324
Integrated LR test	132.940***	69.590 ***	126.620***	69.070***
	[0.000]	[0.000]	[0.000]	[0.000]
Penalized LR test	1574.320***	278.260***	1676.340***	351.060***
	[0.000]	[0.000]	[0.000]	[0.000]
Psuedo-Deviance	5120.778	5184.130	5127.098	5184.648
AIC	5128.778	5192.130	5135.098	5192.648
AICC	5128.993	5192.345	5135.313	5192.863
BIC	5129.902	5193.254	5136.222	5193.772
Dependence	1.929	0.288	2.094	0.398

Notes: See Table 4.5 for notes.

Table 4.7: Scores and riskiness for nested frailty models: PIIGS versus non PIIGS

	County level		D_{BF}			S_{BF}		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
PIIGS	Country Score	Country Riskiness	Group Score	Group Riskiness	Total Riskiness	Group Score	Group Riskiness	Total Riskiness
Portugal	1.231	0.231	1.092	0.092	0.323	1.134	0.134	0.365
Ireland	1.423	0.423	1.092	0.092	0.515	1.134	0.134	0.557
Italy	0.729	-0.271	1.092	0.092	-0.179	1.134	0.134	-0.137
Greece	1.449	0.449	1.092	0.092	0.541	1.134	0.134	0.583
Spain	0.777	-0.223	1.092	0.092	-0.131	1.134	0.134	-0.089
NON-PIIGS								
Austria	1.793	0.793	0.915	-0.085	0.708	0.882	-0.118	0.675
Belgium	0.984	-0.016	0.915	-0.085	-0.101	0.882	-0.118	-0.134
Finland	1.044	0.044	0.915	-0.085	-0.041	0.882	-0.118	-0.074
France	0.583	-0.417	0.915	-0.085	-0.502	0.882	-0.118	-0.535
Germany	0.536	-0.464	0.915	-0.085	-0.549	0.882	-0.118	-0.582
Netherlands	1.209	0.209	0.915	-0.085	0.124	0.882	-0.118	0.091

Notes: Columns 1 and 2 are the country-level scores and riskiness, respectively. Column 2 is obtained by subtracting value 1, as the expected value of the unobserved factors, from the numbers in column 1. The terms D_{PG} (columns 3 to 5) and S_{PG} (columns 6 through 8) are the extraction of risk scores where firms in the PIIGS have similar behaviour of distance to default probability and stock return, respectively. The values of columns 4 and 7 are obtained by subtracting value 1 from columns 3 and 6. The total riskiness for D_{PG} and S_{PG} are constructed by adding columns 2 and 4, and columns 2 and 7.

France, Germany, Italy and Spain. For instance, the risk scores for Portugal and Belgium are 1.231 and 0.984 respectively, values that are shared by firms in these countries. The country riskiness for Portugal and Belgium, 0.231 and -0.016 respectively, are obtained by subtracting the threshold value 1 from their scores. These values imply that firms in Portugal are about 23% more risky as compared to Belgian firms which are about 2% less risky.

The results related to group risk scores are reported in columns 3 and 6 of Table 4.7, and the corresponding values of group riskiness are in columns 4 and 7. Firms in a group are exposed to the same level of risk. For example, firms in the PIIGS group are at least 9% more risky, while those in the non PIIGS are at least 9% less risky. We perform the group score extraction to help determine the total riskiness (the sum of country level riskiness and group-based riskiness), of listed firms in a given country. This tends to offer information on the effects of unobserved factors on firms listed in a country, which is a member of the Euro area and also a member of a group created by the Euro crises. For example, the total riskiness of Belgium is approximately -0.101, whereas Portugal shows a total riskiness score of at least 0.323. These values imply that Belgian firms are at least 10% less risky, while Portuguese firms are at minimum 30% more risky. When comparing the risk levels within the PIIGS, Greek firms seem to be the riskiest, followed by Irish and Portuguese firms, while Spanish and Italian firms are the less risky ones. For the non-PIIGS countries, German firms are the least risky, while the Austrian are the riskiest.

All these result suggests that firms in the Euro area are exposed to both country level and group-based unobserved risk factors. In addition, some countries in the non-PIIGS (PIIGS) group have higher (lower) risk scores and total riskiness. This outcome supports our argument that to measure risk levels of firms accurately, it is fundamental to consider both country and group level. This information may play an important role for financial decision-making process.

In Table 4.8, the results concerning the level of riskiness of firms for the PIISGBF and non PIIGSBF countries are illustrated. Greece is still the country with the highest risk level, followed by Ireland and Portugal, while France has the lowest risk level. For the non PIIGSBF group, Germany is the least risky country, whereas Austria is the riskiest one.

When comparing the results in Table 4.7 and 4.8, the following emerges. First, the riskiness for the PIIGS countries decreases when Belgium and France, with relatively lower risk levels, are included in this group . For example, Portugal's risk level decreases to about 23% in the PIIGSBF group. However, Belgium and France are now 2% and 23% less risky, respectively. This implies that Belgium and France are relatively risky in the PIIGS group than in the non-PIIGS one. Second, the risk levels of non-PIIGSBF group

Table 4.8: Scores and riskiness for nested frailty models: PIIGSBF versus non-PIIGSBF

	County level		D_{BF}			S_{BF}		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
PIIGSBF	Country Score	Country Riskiness	Group Score	Group Riskiness	Total Riskiness	Group Score	Group Riskiness	Total Riskiness
Portugal	1.231	0.231	0.999	-0.001	0.230	1.000	0.000	0.231
Ireland	1.423	0.423	0.999	-0.001	0.423	1.000	0.000	0.423
Italy	0.729	-0.271	0.999	-0.001	-0.272	1.000	0.000	-0.271
Greece	1.449	0.449	0.999	-0.001	0.448	1.000	0.000	0.449
Spain	0.777	-0.223	0.999	-0.001	-0.224	1.000	0.000	-0.223
Belgium	0.984	-0.016	0.999	-0.001	-0.017	1.000	0.000	-0.016
France	0.583	-0.417	0.999	-0.001	-0.418	1.000	0.000	-0.417
NON-PIIGSBF								
Austria	1.793	0.793	1.001	0.001	0.794	1.000	0.000	0.793
Finland	1.044	0.044	1.001	0.001	0.045	1.000	0.000	0.044
Germany	0.536	-0.464	1.001	0.001	-0.463	1.000	0.000	-0.464
Netherlands	1.209	0.209	1.001	0.001	0.210	1.000	0.000	0.209

Notes: The values of columns 4 and 8 are obtained by subtracting value 1 from columns 3 and 7, correspondingly. The total riskiness for D_{BF} and S_{BF} are constructed by adding columns 2 and 4, and columns 2 and 7.

increased, as a result of the absence of Belgium and France in the group. In particular, Germany is between 55-58% less risky, and this range decreases to about 46%. The outcome shows that weak countries in the Euro area tend to benefit, through economic and financial activities, more as compared to those with stronger economies.

In order to ascertain the impact of individual membership of Belgium and France on the PIIGS' group riskiness, we also extract the group score and compute the riskiness for the PIIGSB and PIIGSF groups. The following results are obtained. The risk score falls within the range (1.121, 1.153) and (1.000, 1.003) for PIIGSB and PIIGSF, respectively. Thus, the group riskiness, when Belgium is regarded as a member of the PIIGS, falls within the range 12.10%-15.30%, while that of France is bounded by -0.3% and 0.3%. This seems to suggest that Belgium behaved more like the PIIGS countries than France does, as a result of the crisis.

The above empirical results show that accounting for country level (internal) risk factors may add some explanatory power to default rate models within the Euro area for ranking individual countries in terms of riskiness. However, firms are externally exposed to extra risk induced by the economic and financial activities among the member states of the Euro area, and neglect the potential impacts of group level (external) risk factors on firms' behaviour may likely lead to the underestimation of failure rates and related dependencies among firms.

4.5 Conclusion

In this chapter we employ a mixed effects Cox model that accounts for nested unobserved factors to investigate the hazard rates and dependence structures of public listed firms of the stock exchanges in 11 Euro countries. We apply non-frailty, non-nested frailty and nested frailty models, and employ covariates largely used in the empirical literature, such as distance to default probability, one year trailing stock return, one year trailing market return, firm age, and 3 month T-bill rate. We compare the estimates of the non-frailty model with those of the non-nested frailty model in order to show the importance of accounting for frailty factors at the country level, especially during distressed market periods. Second, we estimate the nested frailty model and compare its performance with that of the non-nested model to show the importance of frailty factors. The analysis is carried out at country (separate entities within the Euro area) and group level. In particular, four different groups of countries are considered: PIIGS, PIIGSBF (Belgium and France are included in the PIIGS group), PIIGSB (Belgium is included in the PIIGS), and PIIGSF (France is regarded as a PIIGS country). Firms are assumed to be exposed to both country-specific and group level unobserved factors. The sample is comprised of 1,422 firms: 905 active firms, 398 failed firms and 119 acquired or merged firms.

The empirical analysis delivers three main results. First, when considering countries as separate entities, country level unobserved factors play an important role in explaining failure rates. In addition, a rise in the distance to default probability causes a decrease in the firms' expected time to default, stocks of firms tend to be better off when the overall economic performance improves, and older firms with high stock returns are less likely to fail. These reveal that, when the effects of unobserved factors for the firms in the Euro area are properly account for, changes in firms' characteristics over a given time horizon are indicative of movements either towards or away from a failure point.

Second, the estimates of the covariates in nested models do not differ from those in the non nested models. However, nested frailty models are able to accurately capture correlations induced by both internal and external factors, and therefore are likely to estimate failure rates with smaller margin of error. This implies that the effect of the crises on firms' behaviour in the Euro area can be better explained by using nested models.

Third, models that do not feature the distance to default probability covariate tend to perform poorly as compared to their counterparts. Thus, it is important to account for changes in firms' leverage in order to accurately gauge the failure rate of firms in the Eurozone.

Chapter 5

Concluding remarks

Probability of default is an extremely important input of risk management. The estimation and prediction of the probability of default rates has gained much attention among investors, regulators, and academics over the last three decades. The probability of default could be used (e.g. Shumway 2001; Duffie et al, 2007, Duan et al., 2012): (i) as an input under the Basel II accord to determine the minimum capital that banks are required to hold; (ii) by financial institutions (especially banks) to discriminate good credit applicants from the bad ones; (iii) by rating agencies to rate firms; and (iv) by academics to test various hypothesis on corporate default such as the impact of failure rates on loan supply.

In order to estimate probability of default, several studies have used risk models that are broadly grouped into two categories, namely the structural and reduced-form models. Structural models consider the changes in a firm's value (assets) over time, where a firm is assumed to be hit by a failure event when it value falls considerably below its liabilities, so that it cannot meet it future obligations (Merton, 1974; Figlewski et al., 2012, page 88); while the reduced form models "treat default as a random event that can strike any firm at any time" (see Figlewski et al., 2012). More specifically, the role that information plays in modelling default risk differentiates between the two types of models (Duffie and Singleton, 1999; Duffie and Lando, 2001; Jarrow and Protter; 2004; Giesecke, 2006, among others). However, reduced form approaches have received more attention than the structural ones, since these models are primarily based on the information available to the market.

Some of the reduced form models used in previous studies assumed that default rate of a firm is unlike to affect the performance of other firms and, even if there exists correlation among firm failures, this can be captured by covariates in the models (see Shumway, 2001; Kavvathas, 2001; Chava and Jarrow, 2004; Charitou et al., 2004; Hillegeist et al., 2004; Duffie et al., 2007; and others). However, the assumption of independence is likely to be unrealistic during unfavourable market conditions, and Das et al. (2007) argued that models which assume the hypothesis of independence are likely to produce less accurate estimates of hazard rates. Consequently, several studies (Duffie et al., 2009; Chava et al., 2011; Koopman et al., 2011; Koopman et al., 2012; Qi et al., 2014; Atsu and Costantini, 2015; Azizpour et al., 2015, among others) have accounted for the impact of unobserved factors on default rates by using frailty factors. These factors are defined as "a random component designed to account for variability due to unobserved individual-level factors that is otherwise unaccounted for by the other predictors in the model" (Kleinbaum and Klein, 2012, page 326).

This thesis explores the impact of unobserved industry level factors taking into account two different market regimes, namely normal and distressed market periods. Also, it examines the impact of informative firm censoring on the failure rates of active firms, whilst paying attention to sector levels unobserved factors. Further, the thesis examines the dynamics of failure rates driven by the exposure of firms to country-based and bloc level unobserved factors. This study uses non-frailty, multivariate non-nested, and nested frailty models in a reduced form, and draws on the assumption of dependence of failure rates, with a particular emphasis on firms' exposure to unobserved risk factors.

Chapter 2 investigates the sector level failure rates and the related dependencies that are induced by sector level unobserved factors. The impact of these factors on failure rates are much higher during distressed market periods. In order to capture this impact more accurately, an additive lognormal frailty model is proposed. The empirical analysis is carried out using public listed firms across various sectors on the London Stock Exchange over the period 1985-2012. We use a set of covariates, namely, distance to default probability (Bharath and Shumway, 2008), stock return, market return, firm age (Shumway, 2001) and 3 month T-bill rate (Das et al, 2007 and Duffie et al., 2007). The study also considers two market regimes, namely normal and distressed market periods.

The empirical results show that, as the market conditions become severe, the adjustment factor increases, thus reflecting the riskiness of active firms, and the effect of distance to default probability averagely reduced. As for the covariates, distance to default probability and market return are positive and statistically significant, while stock return, 3 month T-bill rate and firm age have a negative and significant impact on hazard rates.

In our analysis, we also compare the performance of the additive lognormal frailty and multiplicative gamma frailty models of Atsu and Costantini (2015) and Chava et al. (2011) in terms of: (i) in-and out-samples estimation of sector level failure rates; (ii) overall goodness-of-fit. The results reveal that, during distressed market periods, the model of Atsu and Costantini (2015) tends to produce more accurate estimates as compared to those by the model of Chava et al. (2011). The overall results related to the goodness-of-fit measures show that the additive lognormal frailty model fits the data better than the multiplicative gamma frailty model. These findings may offer relevant information on failure rates among firms at sector level on the London Stock Exchange, and they may be used for portfolio decisions and for designing new regulatory requirements or enhancing existing ones.

Chapter 3 examines the impact of non-default firm exit (informative firm censoring) on the performance of active firms. In this chapter, we use the inverse probability of censoring weighted scheme (see Robins, 1993; Robins and Finkelstein, 2000; Scharfstein and Robins, 2002) in order to correct the potential bias in the estimates of the default rate using the multiplicative gamma frailty model of Chava et al. (2011) and the additive lognormal frailty model of Atsu and Costantini (2015). While these two models offer some flexibility to account for unobserved factors and distressed market periods, they do not account for firm censoring, which is likely to convey relevant information to the market participants. For the analysis, we consider two types of weighting types, the classical and the industry level adjusted schemes. First, classical weights are constructed using only macroeconomic and firm-specific covariates, and it is assumed that the dynamics for firm censoring is the same across all the industries. Second, for the industry level weighted type, we combine macroeconomic and firm-specific covariates with industry level unobserved factors that vary with economic cycles, since firms may leave the market through mergers and acquisitions, which tend to cluster by industry (see e.g. Andrade et al., 2001; Harford, 2005). Thus, firm censoring may vary across industries.

In the empirical analysis, quarterly data from three exchanges in the US, namely NYSE, NASDAQ and NYSE MKT LLC, over the period 1980 - 2013 are used. We employ distance to default probability, market return, stock return, 3 month T-bill rate and industry distress indicator. The empirical results show that the distance to default probability has a higher explanatory power of hazard rates, market return has a positive impact on hazard rate, and a rise in stock return, and 3 month T-bill rate cause a decrease in hazard rates. Further, the bias-corrected models produce more accurate estimates of the scale factor as compared to those of the benchmarks models, and much smaller estimated standard errors, which suggest a relatively larger efficiency.

Chapter 4 examines the impact of default dependence, induced by both country-based (internal) and group level(external) unobserved factors, on default rates in the Euro area over the period 1994Q1-2014Q4. The analysis is conducted at country and group levels, while previous studies have treated countries as standalone (independent) entities, and the potential effects of unobserved factors at a group level are discarded. In particular, we consider the PIIGS and non PIIGS countries along with three extra groups that are formed by including: only Belgium in the PIIGS (PIIGSB); only France in the PIIGS (PIIGSF); both Belgium and France (PIIGSBF) (see also Giordano et al., 2013). This allows us to establish to what extent the crises in the Euro area really affected Belgian and French firms performance along with that of those firms in the PIIGS.

A mixed effects Cox model, which allows the nesting of unobserved factors hierarchically, is used to analyse failure dependence. Quarterly data of listed firms in 11 countries of the Euro area, namely Austria, Belgium, Finland, France, Germany, Greece, Ireland, Italy, Netherlands, Portugal, and Spain, are used. As for the regression analysis, we use the following covariates: distance to default probability, one year trailing stock, one year trailing market return, firm age, and 3 month T-bill rate. The results evidence that: (i) when considering countries as independent entities, unobserved factors at country level impact on hazard rate of firms, a rise in the distance to default probability decreases the firms' expected time to default, whilst market return, stock return, and firm age are negative and statistically significant, and 3 month T-bill rate plays no significant role in the hazard rate specification; (ii) when countries are treated as dependent entities, by adjusting for economic and financial similarities, the effects of the covariates are similar to those in case the hypothesis of independence is assumed, but measures of failure dependence have increased considerably as expected, since unobserved factors at internal and external levels are appropriately captured in the analysis.

Bibliography

Acharya, V. V., Bharath, S. T. & Srinivasan, A. (2007), 'Does industry-wide distress affect defaulted firms? evidence from creditor recoveries', *Journal of Financial Economics* **85**(3), 787–821.

Acharya, V. V., Eisert, T., Eufinger, C. & Hirsch, C. W. (2015), 'Real effects of the sovereign debt crisis in europe: Evidence from syndicated loans', *Available at SSRN 2612855* .

Aitken, N. D. (1973), 'The effect of the eec and efta on european trade: A temporal cross-section analysis', *The American Economic Review* **63**(5), 881–892.

Aldrich, H. & Ruef, M. (2006), *Organizations evolving*, Sage.

Alexandridis, G., Petmezas, D. & Travlos, N. G. (2010), 'Gains from mergers and acquisitions around the world: New evidence', *Financial Management* **39**(4), 1671–1695.

Allison, P. D. (2010), *Survival analysis using SAS: A practical guide*, SAS Institute.

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

Altman, E. I., Marco, G. & Varetto, F. (1994), 'Corporate distress diagnosis: Comparisons using linear discriminant analysis and neural networks (the italian experience)', *Journal of Banking & Finance* **18**(3), 505–529.

Altman, E. I. & Narayanan, P. (1997), 'An international survey of business failure classification models', *Financial Markets, Institutions & Instruments* **6**(2), 1–57.

Andersen, P. K. & Gill, R. D. (1982), 'Cox's regression model for counting processes: a large sample study', *The Annals of Statistics* pp. 1100–1120.

Andrade, G., Mitchell, M. & Stafford, E. (2001), 'New evidence and perspectives on mergers', *Journal of Economic Perspectives* **15**(2), 103–120.

Arghyrou, M. G. & Kontonikas, A. (2012), ‘The emu sovereign-debt crisis: Fundamentals, expectations and contagion’, *Journal of International Financial Markets, Institutions and Money* **22**(4), 658–677.

Atsu, F. & Costantini, M. (2015), ‘Modelling corporate failure dependence of UK public listed firms’.

Azizpour, S., Giesecke, K. & Schwenkler, G. (2015), ‘Exploring the sources of default clustering’, *Available at SSRN 1127792* .

Bak, H. & Maciejewski, S. (2015), ‘Endogeneity and specialization in the european monetary union’, *International Journal of Management and Economics* **46**(1), 7–40.

Baker, M., Pan, X. & Wurgler, J. (2012), ‘The effect of reference point prices on mergers and acquisitions’, *Journal of Financial Economics* **106**(1), 49–71.

Baldwin, R. (1994), ‘Towards an integrated europe, cepr, london’, *Manuscript freely downloadable from http://heiwww. unige. ch/˜ baldwin* .

Baldwin, R. E. & Krugman, P. (2004), ‘Agglomeration, integration and tax harmonisation’, *European Economic Review* **48**(1), 1–23.

Basel Committee on Banking Supervision and Bank for International Settlements (1999), *Credit risk modelling: Current practices and applications*, The Committee.

Baum, J. (1996), ‘Organizational ecology’, *Organizational ecology. In: Clegg, S., Hardy, C., Nord, W.R. (Eds.), Handb. Organ. Stud. Sage, London* pp. 77–114.

Beaver, W. H. (1967), ‘Financial ratios predictors of failure. empirical research in accounting: Selected studies 1966, supplement to vol. 5’, *Journal of Accounting Research* pp. 71–111.

Beaver, W. H. (1968), ‘Market prices, financial ratios, and the prediction of failure’, *Journal of Accounting Research* pp. 179–192.

Bharath, S. T. & Shumway, T. (2008), ‘Forecasting default with the merton distance to default model’, *Review of Financial Studies* **21**(3), 1339–1369.

Bhattacharjee, A., Higson, C., Holly, S. & Kattuman, P. (2009), ‘Macroeconomic instability and business exit: Determinants of failures and acquisitions of uk firms’, *Economica* **76**(301), 108–131.

Bhojraj, S., Lee, C. & Oler, D. K. (2003), 'What's my line? a comparison of industry classification schemes for capital market research', *Journal of Accounting Research* **41**(5), 745–774.

Bonfim, D. (2009), 'Credit risk drivers: Evaluating the contribution of firm level information and of macroeconomic dynamics', *Journal of Banking & Finance* **33**(2), 281–299.

Campbell, J. Y., Hilscher, J. & Szilagyi, J. (2008), 'In search of distress risk', *Journal of Finance* **63**(6), 2899–2939.

Cappelen, A., Castellacci, F., Fagerberg, J. & Verspagen, B. (2003), 'The impact of eu regional support on growth and convergence in the european union', *Journal of Common Market Studies* **41**, 621–644.

Charitou, A., Neophytou, E. & Charalambous, C. (2004), 'Predicting corporate failure: empirical evidence for the uk', *European Accounting Review* **13**(3), 465–497.

Chava, S. & Jarrow, R. A. (2004), 'Bankruptcy prediction with industry effects', *Review of Finance* **8**(4), 537–569.

Chava, S., Stefanescu, C. & Turnbull, S. (2011), 'Modeling the loss distribution', *Management Science* **57**(7), 1267–1287.

Chen, H. (2010), 'Macroeconomic conditions and the puzzles of credit spreads and capital structure', *Journal of Finance* **65**(6), 2171–2212.

Christidis, A. & Gregory, A. (2010), 'Some new models for financial distress prediction in the uk', *Xfi-Centre for Finance and Investment Discussion Paper* (10).

Ciampi, F. (2015), 'Corporate governance characteristics and default prediction modeling for small enterprises. an empirical analysis of italian firms', *Journal of Business Research* **68**(5), 1012–1025.

Clayton, D. G. (1978), 'A model for association in bivariate life tables and its application in epidemiological studies of familial tendency in chronic disease incidence', *Biometrika* **65**(1), 141–151.

Cox, D. R. (1972), 'Regression models and life-tables', *Journal of the Royal Statistical Society. Series B (Methodological)* pp. 187–220.

Cox, D. R. (1975), 'Partial likelihood', *Biometrika* **62**(2), 269–276.

Crawford, J.-A., Fiorentino, R. V. et al. (2005), *The changing landscape of regional trade agreements*, World Trade Organization Geneva.

Das, S. R., Duffie, D., Kapadia, N. & Saita, L. (2007), 'Common failings: How corporate defaults are correlated', *Journal of Finance* **62**(1), 93–117.

De Andrés, J., Landajo, M. & Lorca, P. (2005), 'Forecasting business profitability by using classification techniques: A comparative analysis based on a spanish case', *European Journal of Operational Research* **167**(2), 518–542.

De Grauwe, P. (2010), 'Crisis in the eurozone and how to deal with it', *CEPS Policy Brief* (204).

Deakin, E. B. (1972), 'A discriminant analysis of predictors of business failure', *Journal of Accounting Research* pp. 167–179.

Dempster, A. P., Laird, N. M. & Rubin, D. B. (1977), 'Maximum likelihood from incomplete data via the em algorithm', *Journal of the Royal Statistical Society. Series B (methodological)* pp. 1–38.

Dewaelheyns, N. & Van Hulle, C. (2006), 'Corporate failure prediction modeling: Distorted by business groups' internal capital markets?', *Journal of Business Finance & Accounting* **33**(5-6), 909–931.

Dionne, G. & Laajimi, S. (2012), 'On the determinants of the implied default barrier', *Journal of Empirical Finance* **19**(3), 395–408.

Draper, P. & Paudyal, K. (2006), 'Acquisitions: private versus public', *European Financial Management* **12**(1), 57–80.

du Jardin, P. (2015), 'Bankruptcy prediction using terminal failure processes', *European Journal of Operational Research* **242**(1), 286–303.

du Jardin, P. & Séverin, E. (2012), 'Forecasting financial failure using a kohonen map: A comparative study to improve model stability over time', *European Journal of Operational Research* **221**(2), 378–396.

Duan, J. C., Sun, J. & Wang, T. (2012), 'Multiperiod corporate default predictiona forward intensity approach', *Journal of Econometrics* **170**(1), 191–209.

Duchateau, L. & Janssen, P. (2008), *The frailty model*, Springer Science & Business Media.

Duchateau, L., Janssen, P., Lindsey, P., Legrand, C., Nguti, R. & Sylvester, R. (2002), 'The shared frailty model and the power for heterogeneity tests in multicenter trials', *Computational Statistics & Data Analysis* **40**(3), 603–620.

Duffie, D. (2011), *Measuring corporate default risk*, Oxford University Press.

Duffie, D., Eckner, A., Horel, G. & Saita, L. (2009), 'Frailty correlated default', *Journal of Finance* **64**(5), 2089–2123.

Duffie, D. & Lando, D. (2001), 'Term structures of credit spreads with incomplete accounting information', *Econometrica* **69**(3), 633–664.

Duffie, D., Saita, L. & Wang, K. (2007), 'Multi-period corporate default prediction with stochastic covariates', *Journal of Financial Economics* **83**(3), 635–665.

Duffie, D. & Singleton, K. J. (1999), 'Modeling term structures of defaultable bonds', *Review of Financial studies* **12**(4), 687–720.

Efron, B. (1977), 'The efficiency of cox's likelihood function for censored data', *Journal of the American Statistical Association* **72**(359), 557–565.

Eisenbeis, R. A. (1977), 'Pitfalls in the application of discriminant analysis in business, finance, and economics', *Journal of Finance* **32**(3), 875–900.

El-Agraa, A. M. (2011), *The European Union: economics and policies*, Cambridge University Press.

Fama, E. F. & French, K. R. (1997), 'Industry costs of equity', *Journal of Financial Economics* **43**(2), 153–193.

Farole, T., Rodríguez-Pose, A. & Storper, M. (2011), 'Cohesion policy in the european union: growth, geography, institutions', *JCMS: Journal of Common Market Studies* **49**(5), 1089–1111.

Figlewski, S., Frydman, H. & Liang, W. (2012), 'Modeling the effect of macroeconomic factors on corporate default and credit rating transitions', *International Review of Economics & Finance* **21**(1), 87–105.

Fiorentino, R. V., Verdeja, L. & Toqueboeuf, C. (2007), *The changing landscape of regional trade agreements: 2006 update*, number 12, WTO Discussion Paper.

Flam, H. (1992), 'Product markets and 1992: Full integration, large gains?', *Journal of Economic Perspectives* **6**(4), 7–30.

Gentry, J. A., Newbold, P. & Whitford, D. T. (1987), ‘Funds flow components, financial ratios, and bankruptcy’, *Journal of Business Finance & Accounting* **14**(4), 595–606.

George, G. (2005), ‘Slack resources and the performance of privately held firms’, *Academy of Management Journal* **48**(4), 661–676.

Giesecke, K. (2006), ‘Default and information’, *Journal of Economic Dynamics and Control* **30**(11), 2281–2303.

Gilbert, L. R., Menon, K. & Schwartz, K. B. (1990), ‘Predicting bankruptcy for firms in financial distress’, *Journal of Business Finance & Accounting* **17**(1), 161–171.

Gilson, S. C., John, K. & Lang, L. H. (1990), ‘Troubled debt restructurings: An empirical study of private reorganization of firms in default’, *Journal of Financial Economics* **27**(2), 315–353.

Giordano, R., Pericoli, M. & Tommasino, P. (2013), ‘Pure or wake-up-call contagion? another look at the emu sovereign debt crisis’, *International Finance* **16**(2), 131–160.

Gong, Y., Baker, T. & Miner, A. (2004), Where do organizational routines come from in new ventures, *in* ‘Presentation at the Academy of Management National Meeting’, pp. 6–11.

Grimwade, N. (2007), ‘Measuring the impact of economic integration’, *A. El-Agraa (Derl.), The European Union Economics and Policies içinde, Londra: Financial Times/Prentice Hall* .

Hanagal, D. D. (2011), *Modeling survival data using frailty models*, CRC Press.

Harford, J. (2005), ‘What drives merger waves?’, *Journal of Financial Economics* **77**(3), 529–560.

Hillegeist, S. A., Keating, E. K., Cram, D. P. & Lundstedt, K. G. (2004), ‘Assessing the probability of bankruptcy’, *Review of Accounting Studies* **9**(1), 5–34.

Horta, I. & Camanho, A. (2013), ‘Company failure prediction in the construction industry’, *Expert Systems with Applications* **40**(16), 6253–6257.

Hostak, P., Lys, T., Yang, Y. G. & Carr, E. (2013), ‘An examination of the impact of the sarbanes–oxley act on the attractiveness of us capital markets for foreign firms’, *Review of Accounting Studies* **18**(2), 522–559.

Hougaard, P. (1984), ‘Life table methods for heterogeneous populations: distributions describing the heterogeneity’, *Biometrika* **71**(1), 75–83.

Hougaard, P. (2000), *Analysis of multivariate survival data*, Springer Science & Business Media.

Jacobson, T., Lindé, J. & Roszbach, K. (2013), 'Firm default and aggregate fluctuations', *Journal of the European Economic Association* **11**(4), 945–972.

Jarrow, R. (2001), 'Default parameter estimation using market prices', *Financial Analysts Journal* **57**(5), 75–92.

Jarrow, R. A. & Protter, P. (2004), 'Structural versus reduced-form models: A new information-based perspective', *Journal of Investment Management* p. 110.

Jarrow, R. A. & Turnbull, S. M. (1995), 'Pricing derivatives on financial securities subject to credit risk', *Journal of Finance* **50**(1), 53–85.

Jarrow, R. & Turnbull, S. (1992), 'Credit risk: Drawing the analogy', *Risk Magazine* **5**(9), 63–70.

Karels, G. V. & Prakash, A. J. (1987), 'Multivariate normality and forecasting of business bankruptcy', *Journal of Business Finance & Accounting* **14**(4), 573–593.

Kaski, S., Sinkkonen, J. & Peltonen, J. (2001), 'Bankruptcy analysis with self-organizing maps in learning metrics', *Neural Networks, IEEE Transactions on* **12**(4), 936–947.

Kavvathas, D. (2001), 'Estimating credit rating transition probabilities for corporate bonds'.

Klein, J. P. & Moeschberger, M. L. (2003), *Survival analysis: techniques for censored and truncated data*, Springer Science & Business Media.

Kleinbaum, D. & Klein, M. (2012), 'Survival analysis. a self-learning approach', *Springer, New York, USA* .

Koopman, S. J., Lucas, A. & Schwaab, B. (2011), 'Modeling frailty-correlated defaults using many macroeconomic covariates', *Journal of Econometrics* **162**(2), 312–325.

Koopman, S. J., Lucas, A. & Schwaab, B. (2012), 'Dynamic factor models with macro, frailty, and industry effects for us default counts: the credit crisis of 2008', *Journal of Business & Economic Statistics* **30**(4), 521–532.

Laitinen, E. K. (2007), 'Classification accuracy and correlation: LDA in failure prediction', *European Journal of Operational Research* **183**(1), 210–225.

Lambrecht, B. M. & Myers, S. C. (2007), 'A theory of takeovers and disinvestment', *Journal of Finance* **62**(2), 809–845.

Lancaster, T. (1990), *The econometric analysis of transition data*, number 17, Cambridge university press.

Lando, D., Medhat, M., Nielsen, M. S. & Nielsen, S. F. (2013), 'Additive intensity regression models in corporate default analysis', *Journal of Financial Econometrics* **11**(3), 443–485.

Lane, W. R., Looney, S. W. & Wansley, J. W. (1986), 'An application of the cox proportional hazards model to bank failure', *Journal of Banking & Finance* **10**(4), 511–531.

Lee, E. T. & Wang, J. (2003), *Statistical methods for survival data analysis*, Vol. 476, John Wiley & Sons.

Ludwig, A. (2014), 'A unified approach to investigate pure and wake-up-call contagion: Evidence from the eurozone's first financial crisis', *Journal of International Money and Finance* **48**, 125–146.

Marosi, A. & Massoud, N. (2008), 'you can enter but you cannot leave: Us securities markets and foreign firms', *Journal of Finance* **63**(5), 2477–2506.

Martinussen, T., Scheike, T. H. & Skovgaard, I. M. (2002), 'Efficient estimation of fixed and time-varying covariate effects in multiplicative intensity models', *Scandinavian Journal of Statistics* **29**(1), 57–74.

McGilchrist, C. A. (1993), 'REML estimation for survival models with frailty', *Biometrics* pp. 221–225.

McGilchrist, C. & Aisbett, C. (1991), 'Regression with frailty in survival analysis', *Biometrics* pp. 461–466.

Merton, R. C. (1974), 'On the pricing of corporate debt: The risk structure of interest rates*', *Journal of Finance* **29**(2), 449–470.

Metiu, N. (2012), 'Sovereign risk contagion in the eurozone', *Economics Letters* **117**(1), 35–38.

Minford, P., Gupta, S., Le, V. P. M., Mahambare, V. & Xu, Y. (2015), *Should Britain leave the EU?: an economic analysis of a troubled relationship*, Edward Elgar Publishing.

Moses, D. & Liao, S. S. (1987), 'On developing models for failure prediction', *Journal of Commercial Bank Lending* **69**(7), 27–38.

MSCI (2014), 'Global industry classification standard: A guide to the gics methodology'.

Nelson, R. & Winter, S. (1982), 'An evolutionary theory of economic change. harvard univers. press'.

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

Opler, T. C. & Titman, S. (1994), 'Financial distress and corporate performance', *Journal of Finance* **49**(3), 1015–1040.

Platt, H. D. & Platt, M. B. (1990), 'Development of a class of stable predictive variables: the case of bankruptcy prediction', *Journal of Business Finance & Accounting* **17**(1), 31–51.

Pour, E. K. & Lasfer, M. (2013), 'Why do companies delist voluntarily from the stock market?', *Journal of Banking & Finance* **37**(12), 4850–4860.

Qi, M., Zhang, X. & Zhao, X. (2014), 'Unobserved systematic risk factor and default prediction', *Journal of Banking & Finance* **49**, 216–227.

Raftery, A. E. (1995), 'Bayesian model selection in social research', *Sociological Methodology* **25**, 111–164.

Ripatti, S. & Palmgren, J. (2000), 'Estimation of multivariate frailty models using penalized partial likelihood', *Biometrics* **56**(4), 1016–1022.

Robins, J. M. (1993), Information recovery and bias adjustment in proportional hazards regression analysis of randomized trials using surrogate markers, *in* 'Proceedings of the Biopharmaceutical Section, American Statistical Association', Vol. 24, American Statistical Association, p. 3.

Robins, J. M. & Finkelstein, D. M. (2000), 'Correcting for noncompliance and dependent censoring in an aids clinical trial with inverse probability of censoring weighted (ipcw) log-rank tests', *Biometrics* pp. 779–788.

Rose, A. K. (1999), One money, one market: Estimating the effect of common currecies on trade. camhrige, ma: National bureau of economic reseach, Technical report, Working Paper 7432.

Rudorfer, G. (1995), Early bankruptcy detection using neural networks, *in* 'ACM SIGAPL APL Quote Quad', Vol. 25, ACM, pp. 171–178.

SAS, I. (2014), 'SAS/STAT® 13.2 users guide', *Cary, North Carolina: SAS Institute Inc* .

Sastry, N. (1997), 'A nested frailty model for survival data, with an application to the study of child survival in northeast brazil', *Journal of the American Statistical Association* **92**(438), 426–435.

Scharfstein, D. O. & Robins, J. M. (2002), 'Estimation of the failure time distribution in the presence of informative censoring', *Biometrika* **89**(3), 617–634.

Shumway, T. (2001), 'Forecasting bankruptcy more accurately: A simple hazard model', *Journal of Business* **74**(1), 101–124.

Singer, J. D. & Willett, J. B. (2003), *Applied longitudinal data analysis: Modeling change and event occurrence*, Oxford university press.

Stinchcombe, A. L. & March, J. (1965), 'Social structure and organizations', *Advances in Strategic Management* **17**, 229–259.

Summers, L. H. (1999), 'Distinguished lecture on economics in government: Reflections on managing global integration (digest summary)', *Journal of Economic Perspectives* **13**(2), 3–18.

Tamari, M. (1966), 'Financial ratios as a means of forecasting bankruptcy', *Management International Review* pp. 15–21.

Tang, D. Y. & Yan, H. (2010), 'Market conditions, default risk and credit spreads', *Journal of Banking & Finance* **34**(4), 743–753.

Therneau, T. M. (2015), *coxme: Mixed Effects Cox Models*. R package version 2.2-5.
URL: *http://CRAN.R-project.org/package=coxme*

Therneau, T. M. & Grambsch, P. M. (2000), *Modeling survival data: extending the Cox model*, Springer Science & Business Media.

Therneau, T. M., Grambsch, P. M. & Pankratz, V. S. (2003), 'Penalized survival models and frailty', *Journal of Computational and Graphical Statistics* **12**(1), 156–175.

Tinbergen, J. (1954), *International Economic Integration: Completely Revised of International Economic Co-operation*, Elsevier.

Tinoco, M. H. & Wilson, N. (2013), 'Financial distress and bankruptcy prediction among listed companies using accounting, market and macroeconomic variables', *International Review of Financial Analysis* **30**, 394–419.

Vassalou, M. & Xing, Y. (2004), 'Default risk in equity returns', *Journal of Finance* **59**(2), 831–868.

Wallrafen, J., Protzel, P. & Popp, H. (1996), Genetically optimized neural network classifiers for bankruptcy prediction-an empirical study, *in* 'System Sciences, 1996., Proceedings of the Twenty-Ninth Hawaii International Conference on,', Vol. 2, IEEE, pp. 419–426.

Wienke, A. (2010), *Frailty models in survival analysis*, CRC Press.

Wiklund, J., Baker, T. & Shepherd, D. (2010), 'The age-effect of financial indicators as buffers against the liability of newness', *Journal of Business Venturing* **25**(4), 423–437.

Xie, Y. (1994), 'Log-multiplicative models for discrete-time, discrete-covariate event-history data', *Sociological methodology* **24**, 301–340.

Yeh, C.-Y., Hsu, J., Wang, K.-L. & Lin, C.-H. (2015), 'Explaining the default risk anomaly by the two-beta model', *Journal of Empirical Finance* **30**, 16–33.

Zavgren, C. (1983), 'The prediction of corporate failure: the state of the art', *Journal of Accounting Literature* **2**(1), 1–38.

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

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

Zopounidis, C. & Doumpos, M. (1999), 'A multicriteria decision aid methodology for sorting decision problems: The case of financial distress', *Computational Economics* **14**(3), 197–218.