Corporate Default Models
Empirical Evidence and Methodological Contributions
Christoffersen, Benjamin

Document Version
Final published version

Publication date:
2019

License
CC BY-NC-ND

Citation for published version (APA):
Christoffersen, B. (2019). *Corporate Default Models: Empirical Evidence and Methodological Contributions*. Copenhagen Business School [Phd]. Ph.d. Serie No. 32.2019

Link to publication in CBS Research Portal

General rights
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy
If you believe that this document breaches copyright please contact us (research.lib@cbs.dk) providing details, and we will remove access to the work immediately and investigate your claim.

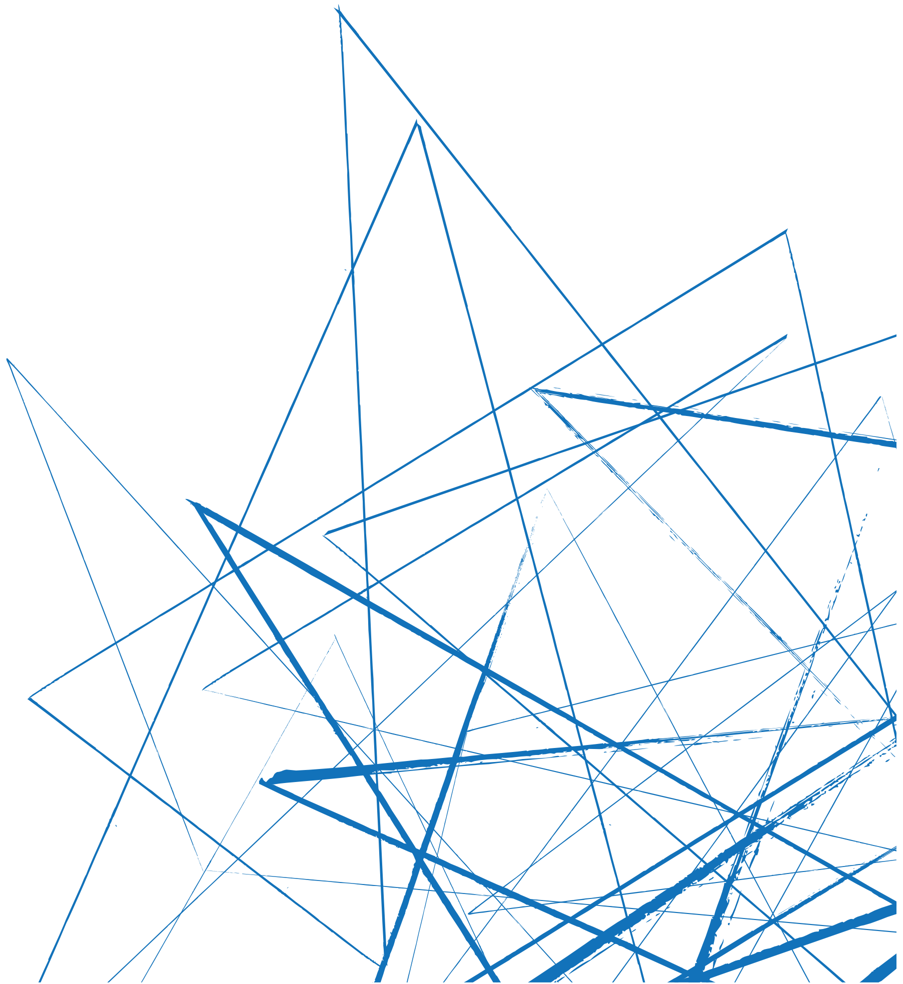
Download date: 30. Mar. 2025

COPENHAGEN BUSINESS SCHOOL
SOLBJERG PLADS 3
DK-2000 FREDERIKSBERG
DANMARK

WWW.CBS.DK

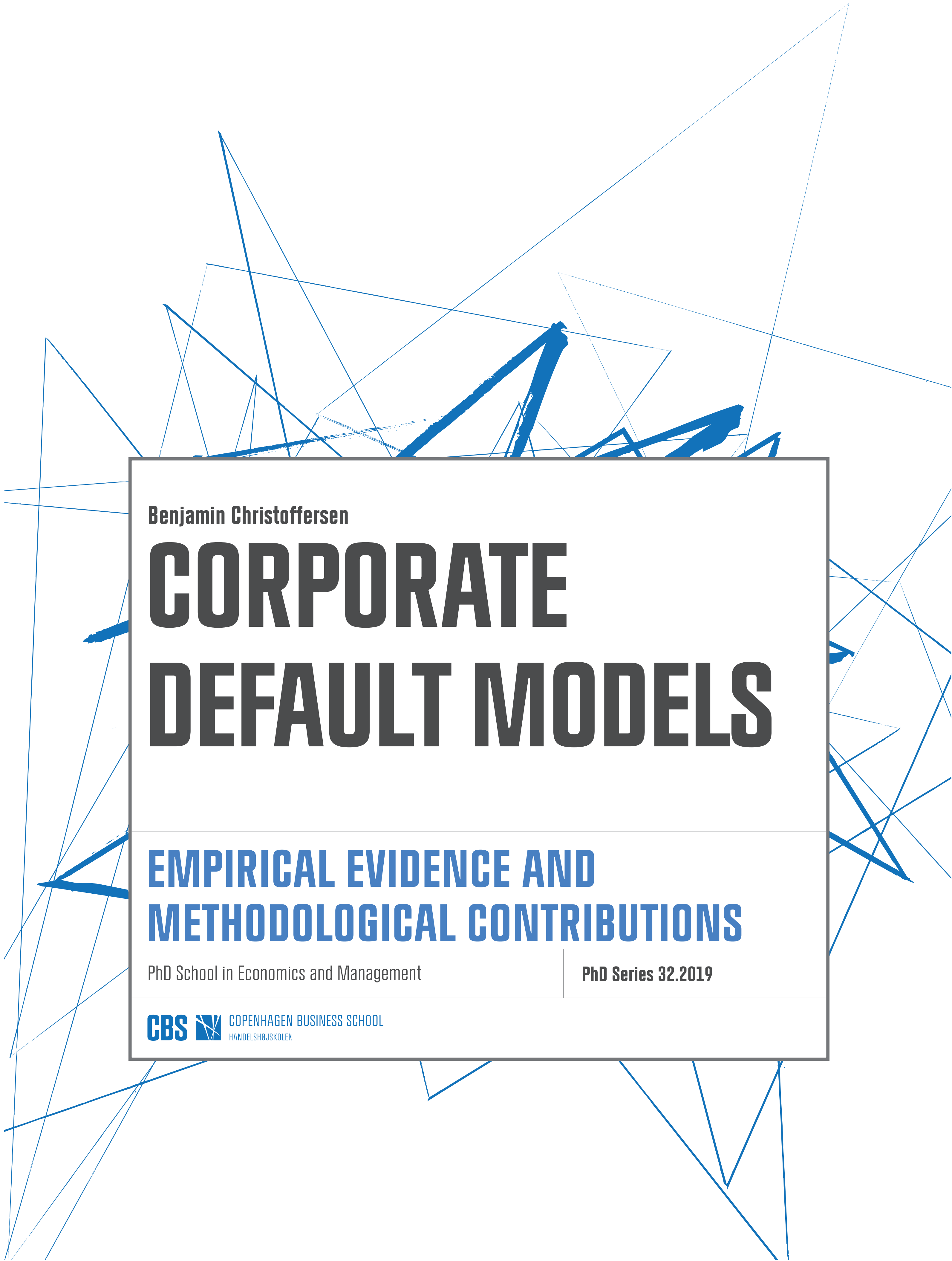
ISSN 0906-6934

Print ISBN: 978-87-93956-06-3
Online ISBN: 978-87-93956-07-0



CORPORATE DEFAULT MODELS: EMPIRICAL EVIDENCE AND METHODOLOGICAL CONTRIBUTIONS

PhD Series 32-2019



Benjamin Christoffersen

CORPORATE DEFAULT MODELS

EMPIRICAL EVIDENCE AND METHODOLOGICAL CONTRIBUTIONS

PhD School in Economics and Management

PhD Series 32.2019

CBS COPENHAGEN BUSINESS SCHOOL
HANDELSHØJSKOLEN

Corporate Default Models

Empirical Evidence and Methodological Contributions

Benjamin Christoffersen

A thesis presented for the degree of
Doctor of Philosophy

Supervisor: Søren Feodor Nielsen
PhD School in Economics and Management
Copenhagen Business School

Benjamin Christoffersen
Corporate Default Models:
Empirical Evidence and Methodological Contributions

1st edition 2019
PhD Series 32.2019

© Benjamin Christoffersen

ISSN 0906-6934
Print ISBN: 978-87-93956-06-3
Online ISBN: 978-87-93956-07-0

The PhD School in Economics and Management is an active national and international research environment at CBS for research degree students who deal with economics and management at business, industry and country level in a theoretical and empirical manner.

All rights reserved.
No parts of this book may be reproduced or transmitted in any form or by any means, electronic or mechanical, including photocopying, recording, or by any information storage or retrieval system, without permission in writing from the publisher.

Preface

This thesis consists of four chapters, all of which are related to credit risk and particularly modeling of default risk. The chapters can be read independently, and the intended audience differs somewhat among them.

The first chapter is methodical; the intended audience consists of statisticians and practitioners who are end users of the software described in the chapter. In particular, the first chapter is written for biostatisticians, statisticians, or practitioners with some prior experience with survival analysis. The chapter shows fast approximate methods to estimate a class hazard models implemented in an open source R package.

The second chapter focuses on default risk models for a broad group of public and private firms. These models are particularly interesting for regulators and banks that wants to evaluate the risk of a corporate debt portfolio with varying exposure. The intended audience consists of academics, particularly those working within finance with default models, as well as practitioners, either on the regulatory or private side. The main question of the chapter is whether the typically observed excess clustering of defaults is due to a misspecification of the dependence between observable variables and the probability of entering into default. While we do find improvements on the firm-level after relaxing standard assumptions, the improvements are substantially smaller than stated previously in the literature. Moreover, we find limited evidence that the more general models fit better on an aggregate scale. Thus, we show an easily implemented random effect model that involves similar relaxations, achieves comparable firm-level performance, and performs better on the aggregate scale.

The third chapter focuses on default models applied exclusively to public firms. Thus, the intended audience is similar to the second chapter. We use a typical data set of U.S. public firms, which has the advantage, compared with the second chapter, that all our firms have market data available. Thus, well-known and powerful market-based predictors of default are available. Unlike in the second chapter, we have substantially fewer firms and observed defaults, which limits the degree to which we can make similar relaxation as in the second chapter. Nevertheless, we still find significant nonlinear effects of some of the variables, some of which are closely related to the Merton model. Moreover, we have substantially more time periods to work with, given that the data set permits us to go down to a monthly instead of yearly scale and we have many more years. The latter allows us to focus more on time-varying effects. We find a time-varying size effect comparable to previous papers but with a model that can be directly used for forecasting. We show that our suggested model performs better out-of-sample on the firm-level and on the aggregate scale during the recent financial crisis.

The last chapter covers details of the particle filtering and smoothing methods implemented in the same package as in the first chapter. The chapter has a more broad intended audience of statisticians and focuses less on the survival analysis applications. All the models and methods are directly applicable to default risk. For example, the methods used in the third chapter are described in full in the last chapter.

Benjamin Christoffersen
Copenhagen, August 2019

Acknowledgments

This thesis would not have been possible without the help and support of several people. Although there are too many to thank, I will emphasize a few people who deserve special recognition.

I started at CBS intending to study finance. Thus, Wharton was the obvious choice for a one-semester exchange during my bachelor's studies. Before the exchange, I was unenthusiastic about streaming a required course on statistical models from CBS where I failed to find a suitable course to use for credit transfer. However, it turned out that while the finance courses I had at Wharton were interesting, I missed out on one of the best courses at CBS. My interest in the course was largely due to Søren Feodor Nielsen, who has since been my supervisor on my bachelor's and master's theses as well as for my Ph.D. project. In fact, this thesis would likely not have been written had it not been for his encouragement, which convinced me that my odds of being accepted as a Ph.D. student were bigger than I had thought.

I already was a bit doubtful about focusing solely on finance since the semester prior to the exchange, when I took Peter Dalgaard's probability theory and statistics course. In the course, we were introduced to what thought at the time was some small project called R, which he is greatly involved in. Little did I know that R would end up being a great interest of mine as well. I am also thankful that he helped me with a visit at Stanford during my Ph.D. studies.

I am grateful for having had David Lando as my secondary supervisor. His expertise in credit risk is only matched by a few people. I appreciate the time he has given me considering all of his obligations.

Thanks to Dorte Kronborg for directing the HA(mat.) and Cand.Merc.(mat.). Given the breadth of the bachelor's program, it yields sufficient information to become interested in a broad range of topics, while the master's program, gives students room to dig into the details. Both programs sparked my curiosity, which contributed to my decision to pursue a Ph.D. Thanks to Niels Richard Hansen for helping me with a visit of almost one year at ETH at the start of my Ph.D. studies. I am thankful for the invitation to visit ETH by Peter Bühlmann and the invitation to visit Stanford by Balasubramanian Narasimhan.

Danmarks Nationalbank turned out to be a central part of my Ph.D. studies. In particular, I would like to thank Thomas Sangill, Pia Mølgaard, and Rastin Matin for the collaboration. My time at the bank has been inspiring, enlightening, and fun. The second and third chapter of this thesis is the product of our collaboration.

I would like to thank everyone at the Department of Finance. I have had an enjoyable and productive time due to everyone at the department. Last, a special thanks goes to my parents and sister, whose seemingly endless support for my academic pursuit is hard to describe. I have tried hard to test their limits, but I have yet to be able to do so.

Introduction and Summaries

Modeling the loss distribution of a corporate debt portfolio is an important task, particularly for regulators and banks. Typically, the main focus is to model the tail of the distributions to ensure the stability of the economy and survival of the bank. This is not an easy task, due to potentially unobservable factors or time-varying associations with either observable macro variables or firm-level variables, combined with often limited observed periods with data; while there are many firms, there is typically substantially less data in the time dimension.

One bottom-up approach to modeling the loss distribution is to decompose the loss to each firm into three components: the exposure at default, the loss given default, and the probability of default. It is clear, though, that because the loss distribution is the sum of losses to each firm, in order to model the loss distribution, one needs to account for, potentially omitted time-varying factors or time-varying effects affecting groups of firms or all firms. More formally, the loss of a debt portfolio in a period $(t-1,t]$ can be decomposed into

$$L_{it} = \sum_{j\in R_{it}} E_{ijt} G_{ijt} Y_{jt}$$

where L_{it} is the loss of portfolio i in interval t, $R_{it} \in \{1,\dots,n\}$ is the set of firms that portfolio i is exposed to in interval t, $E_{ijt} \in (0,\infty)$ is the exposure of portfolio i to firm j at time t, $G_{ijt} \in [0,1]$ is the loss given default of portfolio i to firm j at time t, and $Y_{jt} \in \{0,1\}$ is one if firm i defaults in interval t. Thus, the goal is an accurate model of the joint distribution of $\{E_{ijt}, G_{ijt}, Y_{jt}\}_{j\in R_{it}}$.

As with the majority of the literature, this thesis will focus only on the defaults. Nevertheless, it is clear that to accurately model the loss distribution, L_{it}, and particularly the tail of the loss distribution, one needs an accurate firm-level as well as a joint model of $\{Y_{jt}\}_{j\in R_{it}}$ for fixed t. There is a long history of default models, with noticeable early examples being Beaver (1966) and Altman (1968), which have led to multiple well-documented accounting and market-based predictors of default. However, the joint modeling of defaults is a more recent focus. Highly influential work by Das et al. (2007), Duffie et al. (2009) provide evidence of a shared unobservable effect and a model to account for such an effect. Many papers on more accurate joint models have followed since these publication.

This thesis extends the present literature both with regard to the time-varying and unobservable effects. Moreover, empirical evidence is provided that some assumptions about the association between observed firm-level variables and the probability of default may be violated and that relaxation of these assumptions improves firm-level performance. In particular, the second and third chapters show default models more general than those typically used in the literature. My coauthors and I show that improvements can be made to both the firm-level and the joint distribution of defaults by relaxing typical assumptions such as linearity, additivity, and time-invariant coefficients.

R Packages

The first chapter gives a detailed description of implementations of approximate estimation methods for a class of hazard models that can be used for corporate defaults. Further, the fourth chapter describes the particle-based methods in the same package. This is one of six open source R packages available at the comprehensive R archive network (CRAN) I have authored or coauthored during my studies. Five of the six packages are covered here because only one is described in detail in the chapters of this thesis.

The **pre** package was created by Marjolein Fokkema, and I have subsequently made large contributions to it. The main method in the package is used to derive prediction rule ensembles as suggested by Friedman and Popescu (2008). These rule ensembles typically have the advantage of being easy to interpret while often performing comparably to the best-performing methods. Some of these rule ensembles are a gradient boosted tree model as in Chapter 2, with a subsequent L1 penalized model that includes a term for each node (including nonleaf nodes) in all of the trees. The L1 penalty shrinks many of the coefficients to zero, yielding a sparse solution, typically with substantially fewer terms than used in the original model.

The **DtD** package has a fast C++ implementation to estimate the Merton model (Merton, 1974) from equity and accounting data. Both maximum likelihood estimation and the so-called KMV method are implemented (see e.g., Vassalou and Xing, 2004). The package name is an abbreviation of distance to default, which has been shown to be a powerful predictor of default (see e.g., Duffie et al., 2009, 2007). The package is used to compute the distance to default in Chapter 3.

The **rollRegres** package uses methods from the **LINPACK** library to perform fast rolling and expanding window estimation of linear models. The implementation updates and potentially downdates a QR decomposition, which is both faster and numerically more stable than many other alternatives on CRAN. The idiosyncratic volatility estimates in Chapter 3 are computed with the package.

The **parglm** package contains a parallel implementation of the iterative re-weighted least squares method for generalized linear models. It is possible to use a method in which (a) the inner product of the weighted design matrix is explicitly computed and (b) one where a QR decomposition of the weighted design matrix is computed. The former is faster while the latter is numerically more stable. The package is useful in the scenario where the user does not have access to an optimized **LAPACK** library, **BLAS** library, or a similar linear algebra library. The methods are very similar to those in the `bam` function from the **mgcv** package (Wood et al., 2015) for generalized additive models and the estimation method is written in C++ and computation in parallel is supported with the C++ **thread** library.

The **mssm** package contains similar functionality as one of the **dynamichazard** package's method to approximate the gradient and the method to approximate the observed information matrix. However, the advantage of the **mssm** package is that it allows for more general models, it contains a dual k-d tree approximation like that used in Klaas et al. (2006), and it contains an implementation of two types of antithetic variables like those suggested by Durbin and Koopman (1997). The latter two features are important to obtain a low variance estimate quickly. All the computation is done in C++ and allows for computation in parallel using the C++ **thread** library.

Summaries in English

dynamichazard: Dynamic Hazard Models using State Space Models

The first chapter describes an open source software package containing implementations of fast approximate estimation methods for a class of state space models used in discrete survival analysis that are applicable to corporate default; such applications to corporate default have been featured recently in the literature. The approximation methods are very fast, scale well in both the time-dimension and number of observations, and are easily implemented in parallel. The chapter is intended for the end user of the software and shows how to use the package with a hard disk failure data set.

The implemented approximations are versions of the extended Kalman filter and an unscented Kalman filter. Both methods are widely used in engineering, and the former has connections to the mode approximation technique shown in Durbin and Koopman (2012). The chapter starts with a review of available software to estimate models with time-varying effects in survival analysis, emphasizing whether the models can be used for extrapolation. The latter is important for default models in which the main interest is in the future, where no observations are available for model estimation.

Can Machine Learning Models Capture Correlations in Corporate Distresses?

The second chapter uses a data set of private and public Danish firms. Having a model for both private and public firms is crucial in an economy like the Danish one, because private firms hold a substantial part of the corporate debt. However, market data is not available for private firms, which rules out powerful predictors of default. An advantage of including private firms in the set is that it increases the sample size substantially, potentially allowing for more precise estimates with more complex models.

In this chapter, coauthored with Rastin Matin and Pia Mølgaard, we exploit a large sample of firms to question the typical assumptions of linearity and additivity. We show that adding both nonlinear effects and interactions can provide improvements in the ranking of private and public firms by their default risk. Relaxing the linearity and additivity assumptions also yields more comparable results to those of a commonly used greedy function approximation method. Lastly, none of the models we use that assume independence conditional only on observable covariates yields accurate prediction intervals of the aggregate default rate. Thus, we provide a random effect model that also relaxes the linearity and additivity assumptions. The random effect model provides competitive out-of-sample performance during the most recent period in terms of ranking firms by their default risk, has a significant random effect component, and has wider and more reliable prediction intervals for the industry-wide default rate.

Modeling Frailty Correlated Defaults with Multivariate Latent Factors

The third chapter uses a typical data set from the default literature of U.S. public firms. The data set is long in the time dimension and admits a monthly frequency unlike the data set in the second chapter which is shorter and on an annual frequency. Thus, we are able to potentially pose more questions about the time-varying aspect of the default processes. Further, all firms are public, which allows us to use powerful market-based default predictors. However, there are substantially fewer firms, thus limiting the extent to which we can make similar relaxations as in the second chapter.

In the chapter, joined with Rastin Matin, we show a model where we have relaxed assumptions in previous papers in the literature. In particular, we relax the assumptions of additivity, linearity, and that only the intercept should vary through time. We document evidence of nonlinear effects, an interaction, and a time-varying effect of the relative firm size. An advantage of our model is that it can be directly used for future forecasts. Our final model shows superior out-of-sample ranking of firms by their default risk and an improved forecast of the aggregate default rate in the recent financial crises.

Particle Methods in the dynamichazard Package

The last chapter shows the implemented particle-based methods in the open source software package covered in the first chapter. These methods and the implementation are used in the third chapter to estimate the random effect models. The fourth chapter gives a brief introduction to particle filtering, which is a method to recursively approximate the joint density of an increasing sequence of random variables. Then the implemented particle filters and smoothers are covered along with a Monte Carlo EM-algorithm used to estimate the models and which uses one of the particle smoothers. Finally, the implemented particle methods to approximate the gradient and observed informations are shown.

Summaries in Danish

dynamichazard: Dynamic Hazard Models using State Space Models

Det første kapitel beskriver en open source softwarepakke. Softwarepakken indeholder implementeringer af hurtige approksimations metoder til estimere en klasse state space modeller, som bruges til diskret tids overlevelsesanalyse. Modellerne er direkte anvendelige i kreditrisiko og har været brugt i flere nyere artikler. Approksimationerne er hurtige, skalerer godt i både antallet af observationer og tid, og desuden er det nemt at lave en implementering, som laver udregningerne parallelt. Kapitlet er tiltænkt slutbrugere af softwarepakken, og i kapitlet illustreres anvendelsen af pakken med et harddisk datasæt.

De implementerede approksimationer er en version af et extended Kalman filter og et unscented Kalman filter. Begge metoder har bred anvendelse særlig blandt ingeniører, og den første

har forbindelser til mode approksimationsmetoderne beskrevet i Durbin and Koopman (2012). Kapitlet starter med en gennemgang af eksisterende software til at estimere overlevelsesanalysemodeller med tidsvarierende effekter, med særlig fokus på hvorvidt modellerne kan bruges til fremtidig prognose. Det sidste er særligt vigtigt for modellering af kreditrisiko, da det primære fokus ofte er på fremtidige prognoser, hvor der ikke er nogle observationer tilgængelige, når modellen estimeres.

Can Machine Learning Models Capture Correlations in Corporate Distresses?

I det andet kapitel bruges et datasæt med private og børsnoteret danske virksomheder. Det er vigtigt at have en model for både private og børsnoteret virksomheder i en økonomi som den danske, da en stor del af virksomhedsgælden er til private virksomheder. Mens det ikke er muligt for private virksomheder at bruge markedsbaserede variabler, som har vist sig at være brugbare i fallit sandsynlighedsmodeller, så er fordelen, at der er markant flere virksomheder, hvilket gør det muligt at få mere præcise estimater i komplekse modeller.

Kapitlet er skrevet i samarbejde med Rastin Matin og Pia Mølgaard. I kapitlet undersøger vi betydningen af standardantagelser om linearitet og additivitet. Vi viser, at tilføjelsen af ikke-lineære effekter og inkludering af interaktionseffekter giver en model, der bedre rangerer virksomheder efter deres sandsynlighed for at gå fallit i den følgende periode. Endvidere finder vi, at modellen giver resultater der er tættere på en greedy function approximation metode. Dog kan ingen af modellerne, som antager uafhængighed betinget kun på observerbare variabler give tæt på den nominelle dækningsgrad for prædiktionsintervaller for den aggregeret fallit-sandsynlighedsrate. Derfor estimerer vi en random effekt-model som også inkluderer nogle af de ikke-lineære effekter og interaktionseffekter. Denne model giver sammenlignelige out-of-sample resultater i den seneste periode i vores sample, den har en signifikant random effekt, og den har bredere og tættere på den nominelle dækningsgrad for prædiktionsintervaller for den aggregeret fallit-sandsynlighedsrate .

Modeling Frailty Correlated Defaults with Multivariate Latent Factors

I det tredje kapitel bruger vi et typisk datasæt med amerikanske børsnoteret virksomheder. Datasættet er markant større i tidsdimensionen end i det andet kapitel, da det er en månedlig i stedet for årlig frekvens, samt det indeholder flere år. Derfor kan vi nærmere undersøge tidsvarierende effekter, samt vi kan bruge markedsbaserede variabler, da alle virksomheder er børsnoteret. Vi har dog substantielt færre virksomhed og observerede fallitter, hvorfor vi ikke kan inkludere ikke-lineære effekter og interaktionseffekt i samme grad som i det andet kapitel.

Kapitel er skrevet i samarbejde med Rastin Matin. Vi viser resultater for en model, som inkluderer færre antagelser end typiske modeller i litteraturen. Den sidste model, vi viser, har ikke-lineære effekter, en interaktionseffekt samt mere end kun et tidsvarierende intercept. Særligt viser vi, at der er tegn på en tidsvarierende effekt for den relativ markedstørrelses. En fordel ved modellen vi bruger er, at den kan bruges til prognoser. Vores endelige model er bedre til at rangere virksomheder efter deres fallitsandsynlighed out-of-sample og har mere præcise out-of-sample prædiktioner for den aggregerede fallit-sandsynlighedsrate i den seneste finanskrise.

Particle Methods in the dynamichazard Package

Det sidste kapitel beskriver de implementerede, partikelbaserede metoder i den samme open source softwarepakke som i det første kapitel. Disse metoder bruges til at estimere random effekt-modellerne i det tredje kapitel. Kapitlet starter med en kort introduktion af partikelfiltrering, som er en metode til at lave approksimationer af den simultane fordeling af en sekvens af stokastiske variabler. Derefter bliver de implementerede partikelfiltre og partikel-smoother beskrevet efterfulgt af den brugte Monte Carlo EM-algoritme til at estimere modellerne. Denne algoritme bruger resultatet af en af partikel-smootherne. Sidst gives en beskrivelse af de implementerede metoder til at lave approksimationer af gradienten og den observerede informationsmatrix.

Contents

Preface **i**

Acknowledgments **iii**

Introduction and Summaries **v**

1 dynamichazard: Dynamic Hazard Models using State Space Models **1**
- 1.1 Discrete Hazard Model 7
- 1.2 Methods 9
  - 1.2.1 Extended Kalman Filter 12
  - 1.2.2 Unscented Kalman Filter 16
  - 1.2.3 Sequential Approximation of the Posterior Modes 19
  - 1.2.4 Global Mode Approximation 21
  - 1.2.5 Constant Effects 22
  - 1.2.6 Second Order Random Walk 23
- 1.3 Discrete Versus Continuous Time 24
  - 1.3.1 Continuous Time Model 26
- 1.4 Simulations 28
- 1.5 Comparison with Other Packages 31
- 1.6 Conclusion 33
  - 1.6.1 Further Developments 34

2 Can Machine Learning Models Capture Correlations in Corporate Distresses? **35**
- 2.1 Introduction 36
- 2.2 Related Literature 38
- 2.3 Statistical Models for Predicting Corporate Distress 39
  - 2.3.1 Generalized Linear Models 39
  - 2.3.2 Generalized Additive Models 39
  - 2.3.3 Gradient Tree Boosting 41
  - 2.3.4 Generalized Linear Mixed Models 43
- 2.4 Data and Event Definition 45
  - 2.4.1 Event Definition and Censoring 45
  - 2.4.2 Covariates 46

2.5 Performance of the GLM, the GAM, and the GB Model 47
2.5.1 Evaluating Individual Distress Probabilities 48
2.5.2 Evaluating Aggregated Distress Probabilities 49
2.5.3 Measuring Portfolio Risk Without Frailty 51
2.6 Modeling Frailty in Distresses with a Generalized Linear Mixed Model 52
2.6.1 Predictive Results of the GLMM 54
2.6.2 Frailty Models and Portfolio Risk 55
2.7 Including Macro Variables in the Models 56
2.8 Conclusion 58

Appendices **60**
2.A Variable Selection with Lasso 60
2.B Model Estimation 62
2.B.1 Nonlinear Effects in the GAM and the GLMM 63
2.C Details of the Two Bank Portfolios Example of Section 2.6.2 68
2.D Example of Nonlinear Effect 69

3 Modeling Frailty Correlated Defaults with Multivariate Latent Factors **71**
3.1 Model Specification 73
3.2 Data and Choice of Covariates 75
3.3 Empirical Results 78
3.3.1 Distance to Default 81
3.3.2 Frailty Models 84
3.3.3 Comparison with Other Work 88
3.4 Out-of-sample 90
3.5 Conclusion 91

Appendices **94**
3.A Estimating Frailty Models 94

4 Particle Methods in the dynamichazard Package **97**
4.1 Introduction 98
4.1.1 Overview 99
4.1.2 Methods in the Package 100
4.1.3 Proposal Distributions and Resampling Weights 101
4.2 Nonlinear Conditional Observation Model 109
4.2.1 Where to Make the Expansion 111
4.3 Log-Likelihood Evaluation and Parameter Estimation 111
4.3.1 Vector Autoregression Models 112
4.3.2 Restricted Vector Autoregression Models 112
4.3.3 Estimating Fixed Effect Coefficients 114
4.4 Other Filter and Smoother Options 114

4.5 Generalized Two-Filter Smoother 115
4.6 Gradient and Observed Information Matrix 117

Chapter 1

dynamichazard: Dynamic Hazard Models using State Space Models

Benjamin Christoffersen

Abstract

The **dynamichazard** package implements state space models that can provide a computationally efficient way to model time-varying parameters in survival analysis. I cover the models and some of the estimation methods implemented in **dynamichazard**, apply them to a large data set, and perform a simulation study to illustrate the methods' computation time and performance. One of the methods is compared with other models implemented in R which allow for left-truncation, right-censoring, time-varying covariates, and time-varying parameters.

Keywords: survival analysis, time-varying parameters, extended Kalman filter, EM-algorithm, unscented Kalman filter, parallel computing, R, **Rcpp**, **RcppArmadillo**

Thanks to Hans-Rudolf Künsch and Olivier Wintenberger for constructive conversations. Furthermore, thanks to Søren Feodor Nielsen for feedback and ideas on most aspects of the **dynamichazard** package.

The **dynamichazard** package is for survival analysis with time-varying parameters using state space models. The contribution of this paper is to give an overview of computationally fast nonlinear filtering methods for state space models in survival analysis that scale well in the dimension of the observational equation and illustrate the interface in **dynamichazard** for the methods.

I will start by motivating why one would consider time-varying parameters with the Cox proportional hazards model (Cox, 1972) and give a short overview of available software to estimate time-varying parameters in survival analysis. All mentioned packages or functions are in R (R Core Team, 2018) unless stated otherwise. For simplicity, we start with n individuals where each individual $i = 1, \dots, n$ has a single fixed (not time-varying) covariate x_i and a stop time T_i. Later we look at time-varying multivariate covariate vectors, delayed-entry (also known as left-truncation), and random right-censoring. All three can be handled by the methods in the **dynamichazard** package. Denote the instantaneous hazard rate of an event for individual i at time t by

$$\lambda\left(t \mid x_i\right) = \lim_{h\rightarrow 0^+} \frac{\mathrm{P}\left(t \leq T \leq t + h \mid T \geq t, x_i\right)}{h} \tag{1.1}$$

which can be interpreted as the rate of an event over an infinitesimal unit of time. The Cox proportional hazard model is a commonly used model in survival analysis where the instantaneous hazard rate is

$$\lambda\left(t \mid x_i\right) = \lambda_0(t) \exp\left(\beta x_i\right) \tag{1.2}$$

where $\lambda_0(t)$ is a nonparametric baseline hazard and β is the single parameter in the model. One advantage of the Cox proportional hazard model is the ease of interpreting the parameter: $\exp(\beta) = \lambda\left(t \mid x_i = x' + 1\right) / \lambda\left(t \mid x_i = x'\right)$ is the proportional change of the hazard of a unit increase of the covariate regardless of time, t. However, the effect of a covariate may change across time. For instance, suppose we look at the effect of a drug on the risk of a specific disease and we use age as the time variable. Then different dose levels of a drug may not have the same proportional effect for an adult as for a child.

One way to relax the proportional hazard assumption is to use an interaction between the covariate and a deterministic function of time such that the instantaneous hazard rate is

$$\lambda\left(t \mid x_i\right) = \lambda_0(t) \exp\left(\boldsymbol{\beta}^\top \boldsymbol{g}(t) x_i\right) \tag{1.3}$$

I will refer to the elements of $\boldsymbol{\beta}$ as coefficients and the dot product $\boldsymbol{\beta}^\top \boldsymbol{g}(t)$ as a time-varying parameter to avoid confusion. Thomas and Reyes (2014) show how to estimate the model in Equation (1.3) in R using the `coxph` function in the **survival** package (Terry M. Therneau and Patricia M. Grambsch, 2000, Therneau, 2015), and provide macros to do it in SAS™ (SAS Institute Inc, 2017). It has become even easier with the `coxph` function after Thomas and Reyes' article was published because of the new `tt` argument of `coxph`. Similar functionality is available in Stata (StataCorp, 2017) using the `stcox` command with the `tvc` and `texp` arguments. The model can also be estimated in R with the `cph` function in the **rms** package (Harrell Jr, 2017) and the `coxreg` function in the **eha** package (Broström, 2017).

A downside is that the researcher has to specify the function $\boldsymbol{g}(t)$. A flexible choice is to use a spline such that $\boldsymbol{g}(t) = (g_1(t), g_2(t), \dots, g_k(t))^\top$ where g_is are basis functions. This is done in the **dynsurv** package (Wang et al., 2017) with the `splineCox` function which ultimately uses the `coxph` function in the **survival** package. However, models with several covariates with time-varying effects have a lot of coefficients, and the researcher has to choose the number of knots, and placement of the knots. An alternative for the Cox model is the nonparametric Cox model in the `timecox` function in the **timereg** package (Martinussen and Scheike, 2006). The downside of all the methods is that the researcher has to choose hyperparameters where only some of the implementations provide an automated procedure to select the hyperparameters.

Another option is to use the Aalen's additive regression model (Aalen, 1989) where the instantaneous hazard rate is

$$\lambda\left(t \mid x_i\right) = \lambda_0(t) + \beta(t)x_i \tag{1.4}$$

where $\beta(t)$ is estimated nonparametrically. The Aalen model can be estimated in R with the `aareg` function in the **survival** package, and the `aalen` function in the **timereg** package. The `stlh` command can be used Stata and the **lifelines** package (Davidson-Pilon, 2019) can be used in Python (Rossum, 2017). An issue with the Aalen model is that the estimate of the instantaneous hazard rate can become negative.

A drawback of the nonparametric and semiparametric methods is that they cannot be used to make prediction outside the time range used in the estimation due to the nonparametric parts of the hazard. This is an issue for instance when the objective is to make predictions about the future and we use calendar time as the time scale. One solution is to use a fully parametric function for the cumulative hazard denoted by

$$\Lambda\left(t \mid x_i\right) = \int_0^t \lambda\left(z \mid x_i\right) dz \tag{1.5}$$

In particular, we can model the log cumulative hazard function with a restricted cubic spline for the intercept such that

$$\Lambda\left(t \mid x_i\right) = \exp\left(\boldsymbol{\gamma}^\top \boldsymbol{k}\left(\log(t)\right) + \boldsymbol{\beta}^\top \boldsymbol{g}\left(\log(t)\right) x_i\right) \tag{1.6}$$

where $\boldsymbol{\gamma}$ is a coefficient vector, $\boldsymbol{k}(z)$ is a vector of basis functions, and $\boldsymbol{g}(z)$ is a vector of basis functions to get a time-varying parameter like in Equation (1.3). This is implemented in the **rstpm2** package (Clements and Liu, 2016, Liu et al., 2016, 2017), the **flexsurv** package (Jackson, 2016), and the `stpm2` command (Lambert and Royston, 2009) in Stata. All of the packages can fit other models than generalization like in Equation (1.6) of the proportional hazard model (the special case of Equation (1.6) when $\boldsymbol{g}\left(\log(t)\right)$ is constant). Further, the **rstpm2** package includes penalized methods. This is useful as it allows for flexible splines with a large number of basis functions that do not overfit. The `hare` function in the **polspline** package (Kooperberg, 2015) is another alternative which uses linear splines instead of restricted cubic splines, and models the hazard function such that

$$\lambda\left(t \mid x_i\right) = \exp\left(\boldsymbol{\gamma}^\top \boldsymbol{k}(t) + \boldsymbol{\beta}^\top \boldsymbol{g}(t)x_i\right) \tag{1.7}$$

Another alternative is to consider discrete time hazard models. Let T be the event time and

$$Y_t = \begin{cases} 1 & \text{if } T \in (t-1, t] \\ 0 & \text{otherwise} \end{cases}, \qquad t = 1, 2, \dots \tag{1.8}$$

be an indicator for whether there is an event between time $t-1$ and t. Then we model the conditional probability of event given survival up to time $t-1$ by

$$l\left(\mathrm{P}\left(Y_t = 1 \mid T > t-1\right)\right) = \boldsymbol{\gamma}^\top \boldsymbol{k}\left(t\right) + \boldsymbol{\beta}^\top \boldsymbol{g}\left(t\right) x_i \tag{1.9}$$

where l is a link function and $\boldsymbol{k}(z)$ and $\boldsymbol{g}(z)$ are vectors of basis functions to get a time-varying parameter as before (see Tutz and Schmid, 2016, chapter 5 for examples). Equation (1.9) is the discrete hazard rate on the link scale. Software to penalize the coefficients when, potentially, large dimensional $\boldsymbol{k}$ and $\boldsymbol{g}$ are used are available and well established. As mentioned with the **rstpm2** package, this is important to allow for high dimensional splines that do not overfit. A few examples of packages are **glmnet** (Simon et al., 2011), **glmpath** (Park and Hastie, 2013), **mgcv** (Wood, 2017), and **penalized** (Goeman, 2010). A discrete hazard model is an obvious choice if the outcomes are only observed in discrete time intervals and may yield similar results to a continuous time model if the discrete time periods are sufficiently narrow and the time of events is observed. A nonparametric alternative is the temporal process regression in the **tpr** package (Fine et al., 2004, Yan and Fine, 2004) which uses nonparametric time-varying effects in generalized linear models.

While all of the parametric models allow for extrapolation beyond the observed time period, the values of the prediction depend on the chosen type of splines. Some other survival analysis options in R are the **pch** package (Frumento, 2016) and **eha** package. The **pch** package and the `phreg` function in the **eha** package with argument `dist = 'pch'` fit time-varying parameters by dividing the time into periods $(s_0, s_1], (s_1, s_2], \dots, (s_{d-1}, s_d]$ and using separate coefficients in each interval for time-varying parameters. Thus, instantaneous hazards are piecewise constant. If all parameters are time-varying then the instantaneous hazard rate is

$$\lambda\left(t \mid x_i\right) = \exp\left(\gamma_k + \beta_k x_i\right), \qquad k : s_{k-1} < t \leq s_k \tag{1.10}$$

Similar models are easily estimated with `streg` and `stsplit` commands in Stata or a bit of pre-processing followed by the `transreg` procedure in SAS. The models have a lot of coefficients even with a moderate amount of time periods, and can yield unstable coefficients with large jumps. Forecasting of future outcomes conditional on the covariate are predictions from an exponential distribution for the most recent period, $(s_{d-1}, s_d]$, which may be based on a sparse amount of data. Moreover, the number of points where the coefficients jump, d, and the location of the jumps, $s_1, s_2, \dots, s_d$, have to be chosen. The `bayesCox` function in the **dynsurv** package alleviates these issues in a Bayesian analysis where the number of jumps and location of the jumps is a random variable over a fixed size grid of jump locations, the baseline hazard in each interval is gamma distributed, and the coefficients for the covariates follow a first order random walk. Though, the implemented MCMC method is very computationally expensive.

The dynamichazard package

The **dynamichazard** package adds to the existing literature by providing a simple and efficient implementation of models covered by Fahrmeir (1992, 1994). In the two papers, Fahrmeir shows how to model time-varying parameters by using discrete time state space models where the parameters are assumed to be piecewise constant. One possible model in this framework is a model with instantaneous hazard rate like in the piecewise constant hazard model shown in Equation (1.10) given by

$$\begin{aligned} \lambda\left(t \mid x_i\right) &= \exp\left(\gamma_k + \beta_k x_i\right), \qquad k = \lceil t \rceil \\ \left(\gamma_k, \beta_k\right) &\sim f\left(\gamma_{k-1}, \beta_{k-1}\right) \end{aligned} \tag{1.11}$$

where f is a multivariate normal distribution with a mean depending on γ_{k-1} and β_{k-1}, $\lceil t \rceil$ is the ceiling of t, and we use time periods with length 1. This is implemented in the **dynamichazard** package. An advantage of the state space approach is that the issues with the number of coefficients in the piecewise constant hazard model shown in Equation (1.10) are eased by the dependence induced through f. Further, the state space model provides a parametric model for the parameters that allows one to project future parameter values. Thus, it is easy to make future forecasts, and all available data is used in the estimation. Moreover, the models can be estimated approximately with fast methods with a linear computational cost relative to the number of observed individuals, cubic computational cost relative to the number of parameters, and are easily computed in parallel.

There is a lot of software options for fitting general state space models. Two reviews of packages in R for linear Gaussian models from 2011 are Petris and Petrone (2011) and Tusell (2011). They only briefly mention nonlinear methods. The **KFAS** package (Helske, 2017) is the most closely related package to **dynamichazard**. **KFAS** can be used for survival analysis although this is not the primary focus of the package. The researcher can also estimate the models in **dynamichazard** with software like the **pomp** package (King et al., 2016, 2017) in R, Stan (Team, 2017), the **SSM** toolbox (Peng and Aston, 2011) in MATLAB, the Control System Toolbox™ in MATLAB, the **SsfPack** library (Koopman et al., 2008) in C, the **pyParticleEst** library (Nordh, 2017) in Python, the **vSMC** library (Zhou, 2015) in C++, and the **SMCTC** library (Johansen, 2009) in C++ just to name a few. Because all of these are quite general, using them to set up models like those in **dynamichazard** is cumbersome or computationally expensive. Some are computationally expensive as they are intended for a few outcomes at each time point which is common in the state space model literature. This is not the case for the model given in Equation (1.11) as the number of observations at risk at each point in time may be large.

This package is motivated by Fahrmeir (1992, 1994). The current implementation uses the

EM-algorithm from these papers. The reader may want further information on the filters covered later, as this paper only introduces them briefly. Durbin and Koopman (2012, chapter 4) cover the Kalman filter, which provides a basis for understanding all the filters in the package. Fahrmeir (1992, 1994) covers the extended Kalman filter this package uses, whereas Durbin and Koopman (2012, section 10.2) cover the more common form of the extended Kalman filter. Durbin and Koopman (2012, section 10.3) and Wan and Merwe (2000) provide an introduction to the unscented Kalman filter. Another resource is Hartikainen et al. (2011) who introduce the Kalman filter, extended Kalman filter, and unscented Kalman filter.

The hard disk data set example is mainly chosen because it is publicly available and moderately large. It contains data on the hard disk failure times from the time of installation. The hard disks are from Backblaze which a data storage provider. The hard disk survival time seem to differ substantially between both manufacturers and hard disk versions motivating a different process for each hard disk version.

The typical application of the implemented models are cases where we expect time-varying effects, assume that a model like in Equation (1.11) is a good approximation of the hazard rates, and we are interested in future forecasts. Examples are churn analysis where the rate at which customers leave a company may have non-constant associations with observable variables due to e.g., a competitor who launches a marketing campaign targeted towards a group of customers, or firm default prediction where changes in banks lending behavior may effect the rate at which firms with high debt default. The examples can be modeled with calendar time as the time scale and using delayed entry for customers who join a company's service at different points in time or firms who incorporate at different points in time. Another example is mortality rates in life insurance where the rates may be varying in calendar time.

In all cases, we may be interested in predicting future hazard rates and not present ones. Thus, having a model for the relation between present parameter values and future parameter values is useful. The hard disk data set presented later is similar in the sense that hard disk of the same type are typically installed within a short time period. Thus, we do not have data for a given type of hard disk in the range we are interested as they are all installed at roughly the same time.

All methods are implemented in C++ with use of **BLAS** and **LAPACK** (Anderson et al., 1999) either by direct calls to the methods or through the C++ library **Armadillo** (Sanderson and Curtin, 2016). The implemented estimations methods are fast because the algorithms are fast, the methods are implemented in C++, and most of them support computations in parallel. The reported computational complexities in the rest of the paper are based on a single iteration of the EM-algorithm.

The rest of this paper is organized as follows. Section 1.1 covers the discrete hazard model implemented in the package. Section 1.2 shows the EM-algorithm on which all the methods are based, followed by four different filters used in the E-step. A data set with hard disk failures will be used throughout this section to illustrates how to use the methods. Section 1.3 covers the two implemented models. Section 1.4 illustrate the methods' performance and the computation time of the methods on simulated data. One of the methods is compared with some of the above mentioned methods from other packages in Section 1.5. I conclude and discuss extensions in Section 1.6.

1.1 Discrete Hazard Model

I will start by introducing the discrete time model for survival analysis. Outcomes in the model are binary as for the model in Equation (1.9). Either an individual has an event or not within each interval. I generalize in Section 1.3 to a continuous time model. We are observing individual $1, 2, \dots, n$ who each has an *event* at time $T_1, T_2, \dots, T_n$. Further, we separate the part of the timeline we observe into d equidistant intervals. We define the *left-truncation and right-censoring indicators* $D_{i1}, D_{i2}, \dots, D_{id}$ with $D_{it} \in \{0, 1\}$. The indicator is one if the individual is either left-truncated or right-censored. By definition I set $D_{ik} = 1$ for $k > t$ if we observe an event for individual i at time t. I define the following series of outcome indicators for each individual

$$Y_{it} = 1_{\{T_i \in (t-1,t]\}} = \begin{cases} 1 & \text{if } T_i \in (t-1, t] \\ 0 & \text{otherwise} \end{cases} \tag{1.12}$$

y_{it} denotes whether individual i experiences an event in interval $(t-1, t]$. We observe covariate vector $\boldsymbol{x}_{ij}$ for each individual i if $D_{ij} = 0$ where the latter subscripts correspond to the interval number. Next, the *risk set* in time interval t is given by

$$R_t = \{ i \in \{1, \dots, n\} : D_{it} = 0 \} \tag{1.13}$$

I will refer to this as the *discrete time risk set*, as I will introduce a continuous time version later. The risk of an event for a given individual i in interval t is given by

$$\mathrm{P}\left(Y_{it} = 1 \mid \boldsymbol{\alpha}_t, T_i > t-1\right) = h(\boldsymbol{\alpha}_t^\top \boldsymbol{x}_{it}) \tag{1.14}$$

$\boldsymbol{\alpha}_t$ is the state vector in interval t, and h is the inverse link function. The inverse logit function, $h(\eta) = \exp(\eta)/(1+\exp(\eta))$, is used by default. The model written in the state space form is

$$\begin{array}{lll} \mathrm{E}\left(\boldsymbol{Y}_t \mid \boldsymbol{\alpha}_t\right) = \boldsymbol{z}_t(\boldsymbol{\alpha}_t) & & \\ \boldsymbol{\alpha}_{t+1} = \mathbf{F}\boldsymbol{\alpha}_t + \mathbf{R}\boldsymbol{\eta}_t & \boldsymbol{\eta}_t \sim N(\mathbf{0}, \mathbf{Q}) \quad , & t = 1, \dots, d \\ & \boldsymbol{\alpha}_0 \sim N(\boldsymbol{\mu}_0, \mathbf{Q}_0) & \end{array} \tag{1.15}$$

where $\boldsymbol{Y}_t = (Y_{it})_{i \in R_t}$. Notice that the bold '$\mathbf{R}$' is a system matrix, whereas the italic 'R_t' is a risk set. The equation for $\boldsymbol{Y}_t$ is referred to as the *observational equation*. $\boldsymbol{\alpha}_t$ is the *state vector* with the corresponding *state equation*. Further, I denote the observational equation's conditional covariance matrix by $\mathbf{H}_t(\boldsymbol{\alpha}_t) = \mathrm{Var}\left(\boldsymbol{Y}_t \mid \boldsymbol{\alpha}_t\right)$. The mean $\boldsymbol{z}_t(\boldsymbol{\alpha}_t)$ and variance $\mathbf{H}(\boldsymbol{\alpha}_t)$ are state dependent with

$$\begin{aligned} z_{kt}(\boldsymbol{\alpha}_t) &= \mathrm{E}\left(Y_{i_k t t} \mid \boldsymbol{\alpha}_t\right) = h(\boldsymbol{\alpha}_t^\top \boldsymbol{x}_{i_k t t}) \\ H_{kk't}(\boldsymbol{\alpha}_t) &= \begin{cases} z_{i_k t t}(\boldsymbol{\alpha}_t)(1 - z_{i_k t t}(\boldsymbol{\alpha}_t)) & k = k' \\ 0 & \text{otherwise} \end{cases} \end{aligned} \tag{1.16}$$

where $R_t = \{i_{1t}, \dots, i_{n_t t}\}$. The state equation is implemented with a first and second order random walk. The first order random walk model has $\mathbf{F} = \mathbf{R} = \mathbf{I}_m$ where m is the number of time-varying parameters and $\mathbf{I}_m$ is the identity matrix with dimension m. I let q denote the dimension for the state vector. Thus, $q = m$ for the first order random walk model. For the second order random walk model, we have

$$\mathbf{F} = \begin{pmatrix} 2\mathbf{I}_m & -\mathbf{I}_m \\ \mathbf{I}_m & \mathbf{0}_m \end{pmatrix}, \qquad \mathbf{R} = \begin{pmatrix} \mathbf{I}_m \\ \mathbf{0}_m \end{pmatrix} \tag{1.17}$$

where $\mathbf{0}_m$ is a $m \times m$ matrix with zeroes in all entries. That is, we have taken the difference twice. To see this, let $\boldsymbol{\alpha}_t = (\boldsymbol{\xi}_t^\top, \boldsymbol{\xi}_{t-1}^\top)^\top$. Then Equation (1.17) implies that $\boldsymbol{\xi}_t - 2\boldsymbol{\xi}_{t-1} + \boldsymbol{\xi}_{t-2} = \boldsymbol{\epsilon}_t$ which states that second-order difference are independent normally distributed. We assume throughout the rest of the paper that $\mathbf{R}$ has orthogonal columns ($\mathbf{R}^\top\mathbf{R} = \mathbf{I}_m$). Further, we replace the linear predictor, $\boldsymbol{\alpha}_t^\top \boldsymbol{x}_{i_{kt}t}$, in Equation (1.16) with

$$z_{kt}(\boldsymbol{\alpha}_t) = h(\boldsymbol{\xi}_t^\top \boldsymbol{x}_{i_{kt}t}) \tag{1.18}$$

Notice that the dimension of the state vector is $q = 2m$, which affects the computational complexity. The complete data likelihood of the model can be written as follows by an application of the Markov property of the model

$$L(\mathbf{Q}, \mathbf{Q}_0, \boldsymbol{\mu}_0) = p\left(\boldsymbol{\alpha}_0\right) \prod_{t=1}^{d} p\left(\boldsymbol{\alpha}_t \mid \boldsymbol{\alpha}_{t-1}\right) \prod_{i \in R_t} p\left(y_{it} \mid \boldsymbol{\alpha}_t\right) \tag{1.19}$$

where p denotes (conditional) density functions or probability mass functions. Thus, the log-likelihood (... depends on an omitted normalization constant) is

$$\begin{aligned} \log L\left(\mathbf{Q}, \mathbf{Q}_0, \boldsymbol{\mu}_0\right) = &-\frac{1}{2}\left(\boldsymbol{\alpha}_0 - \boldsymbol{\mu}_0\right)^\top \mathbf{Q}_0^{-1}\left(\boldsymbol{\alpha}_0 - \boldsymbol{\mu}_0\right) \\ &-\frac{1}{2}\sum_{t=1}^{d}\left(\boldsymbol{\alpha}_t - \mathbf{F}\boldsymbol{\alpha}_{t-1}\right)^\top \mathbf{R}\mathbf{Q}^{-1}\mathbf{R}^\top\left(\boldsymbol{\alpha}_t - \mathbf{F}\boldsymbol{\alpha}_{t-1}\right) \\ &-\frac{1}{2}\log\left|\mathbf{Q}_0\right| - \frac{2}{d}\log\left|\mathbf{Q}\right| \\ &+\sum_{t=1}^{d}\sum_{i \in R_t} l_{it}(\boldsymbol{\alpha}_t) + \dots \end{aligned} \tag{1.20}$$

$$l_{it}(\boldsymbol{\alpha}_t) = y_{it}\log h(\boldsymbol{x}_{it}^\top\boldsymbol{\alpha}_t) + (1 - y_{it})\log\Big(1 - h(\boldsymbol{x}_{it}^\top\boldsymbol{\alpha}_t)\Big) \tag{1.21}$$

This completes the introduction of the discrete time model. I continue with the methods used to fit the model. In the rest of the paper, all comments about computational costs are assuming that the number of observations, n, is much greater than the number of coefficients, q.

1.2 Methods

I will focus on the `ddhazard` function throughout this article. All the methods that are available with this function use the M-step and parts of the E-step of the EM-algorithm described in Fahrmeir (1992, 1994). Moreover, the method is exactly as in the former mentioned papers when the extended Kalman filter with one iteration (which I introduce later) is used. The EM-algorithm is similar to the method in Shumway and Stoffer (1982) but with a nonlinear observational equation. The unknown hyperparameters in Equation (1.15) are the covariance matrices $\mathbf{Q}$ and $\mathbf{Q}_0$ and the initial state mean $\boldsymbol{\mu}_0$. $\mathbf{Q}$ and $\boldsymbol{\mu}_0$ will be estimated in the M-step of the EM-algorithm. It is common practice with Kalman filters to set the diagonal elements of $\mathbf{Q}_0$ to large fixed values such that one term is removed from Equation (1.20). I use the following notation for the conditional mean and covariance matrix

$$\boldsymbol{a}_{t|s} = \mathrm{E}\left(\boldsymbol{\alpha}_t \mid \boldsymbol{y}_1, \dots, \boldsymbol{y}_s\right), \qquad \mathbf{V}_{t|s} = \mathrm{Var}\left(\boldsymbol{\alpha}_t \mid \boldsymbol{y}_1, \dots, \boldsymbol{y}_s\right) \tag{1.22}$$

Notice that the letter ‘a’ is used for the mean estimates, whereas ‘alpha’ is used for the unknown states. The notation both covers filter estimates in the case where $s \leq t$ and smoothed estimates when $s > t$. I suppress the dependence on the covariates, $\boldsymbol{x}_{it}$, to simplify the notation.

The EM-algorithm is shown in Algorithm 1. The matrices $\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_d$ are the design matrices given by the risk sets $R_1, R_2, \dots, R_d$ and the covariate vectors. The only unspecified part is the filter in line 4 of Algorithm 1. Notice that the other lines involve only products of matrices and vectors of dimension equal to the state space vector's dimension, q. Moreover, the computational cost is independent of the size of the risk sets for the specified parts of Algorithm 1. Thus, the computational complexity so far is $\mathcal{O}\left(q^3 d\right)$, where d is the number of intervals. The threshold for convergence is determined by the `eps` of the `ddhazard_control` functions which is passed as the `control` argument to `ddhazard` (e.g., `ddhazard_control(eps = 0.001, ...)`) similar to the `glm` function. The EM-algorithm tends to converge slowly toward the end. The filters implemented for line 4 of Algorithm 1 are an extended Kalman filter (EKF), an unscented Kalman filter (UKF), a sequential mode approximation (SMA), and a global mode approximation (GMA). I will cover these in their respective order. First, I will briefly cover the Kalman filter and use the Kalman filter to illustrate why we need to use approximations in the filters. Then, I will give a brief overview of all the methods before covering each method in more detail.

The Kalman filter can be applied in line 4 of Algorithm 1 when the observed outcomes, $\boldsymbol{y}_t$, in Equation (1.15) are normally distributed conditional on the state vector and depend linearly on the state vector. The Kalman filter is a two-step recursive algorithm. The first step in the Kalman Filter is the *prediction step* where we estimate $\boldsymbol{a}_{t|t-1}$ and $\mathbf{V}_{t|t-1}$ based on $\boldsymbol{a}_{t-1|t-1}$ and $\mathbf{V}_{t-1|t-1}$. Secondly, we carry out the *correction step* where we estimate $\boldsymbol{a}_{t|t}$ and $\mathbf{V}_{t|t}$ based on $\boldsymbol{a}_{t|t-1}$ and $\mathbf{V}_{t|t-1}$ and the observations. We repeat the process until $t = d$. One advantage with Kalman filter is that both steps can be solved analytically. However, there is no analytical solution for the models covered in this paper. While the prediction step can be solved analytically as the state model is linear and Gaussian, the correction step cannot because of the non-Gaussian distribution of the outcomes given the state vector. Thus, an approximation is needed. The **ddhazard** function

Algorithm 1 EM algorithm with unspecified filter. $\|\cdot\|_2$ is the L2 norm.

```
Input:
    $\mathbf{Q}, \mathbf{Q}_0, \boldsymbol{\mu}_0, \mathbf{X}_1, \dots, \mathbf{X}_d, \boldsymbol{y}_1, \dots, \boldsymbol{y}_d, R_1, \dots, R_d$
    Convergence threshold $\epsilon$
1:  Set $\boldsymbol{a}^{(0)}_{0|0} = \boldsymbol{\mu}_0$ and $\mathbf{Q}^{(0)} = \mathbf{Q}$
2:  for $k = 1, 2, \dots$ do
3:      procedure E-STEP
4:          Apply filter with $\boldsymbol{a}^{(k-1)}_{0|0}$, $\mathbf{Q}^{(k-1)}$ and $\mathbf{Q}_0$ to get
                $\boldsymbol{a}_{1|0}, \boldsymbol{a}_{1|1}, \boldsymbol{a}_{2|1}, \dots, \boldsymbol{a}_{d|d-1}, \boldsymbol{a}_{d|d}$ and
                $\mathbf{V}_{1|0}, \mathbf{V}_{1|1}, \mathbf{V}_{2|1}, \dots, \mathbf{V}_{d|d-1}, \mathbf{V}_{d|d}$
            Apply smoother by computing
5:          for $t = d, d-1, \dots, 1$ do
6:              $\mathbf{B}^{(k)}_t = \mathbf{V}_{t-1|t-1}\mathbf{F}\mathbf{V}^{-1}_{t|t-1}$
7:              $\boldsymbol{a}^{(k)}_{t-1|d} = \boldsymbol{a}_{t-1|t-1} + \mathbf{B}^{(k)}_t(\boldsymbol{a}^{(k)}_{t|d} - \boldsymbol{a}_{t|t-1})$
8:              $\mathbf{V}^{(k)}_{t-1|d} = \mathbf{V}_{t-1|t-1} + \mathbf{B}^{(k)}_t(\mathbf{V}^{(k)}_{t|d} - \mathbf{V}_{t|t-1})\left(\mathbf{B}^{(k)}_t\right)^\top$
9:      procedure M-STEP
            Update the initial state and the covariance matrix by
10:         $\boldsymbol{a}^{(k)}_{0|0} = \boldsymbol{a}^{(k)}_{0|d}$
11:         $\mathbf{Q}^{(k)} = \frac{1}{d}\sum_{t=1}^{d} \mathbf{R}^\top \left( \left(\boldsymbol{a}^{(k)}_{t|d} - \mathbf{F}\boldsymbol{a}^{(k)}_{t-1|d}\right)\left(\boldsymbol{a}^{(k)}_{t|d} - \mathbf{F}\boldsymbol{a}^{(k)}_{t-1|d}\right)^\top + \mathbf{V}^{(k)}_{t|d} - \mathbf{F}\mathbf{B}^{(k)}_t\mathbf{V}^{(k)}_{t|d} - \left(\mathbf{F}\mathbf{B}^{(k)}_t\mathbf{V}^{(k)}_{t|d}\right)^\top + \mathbf{F}\mathbf{V}^{(k)}_{t-1|d}\mathbf{F}^\top \right)\mathbf{R}$
        Stop the if sum of relative norm of changes is below the threshold
12:     $\sum_{t=0}^{d} \frac{\left\|\boldsymbol{a}^{(k)}_{t|d} - \boldsymbol{a}^{(k-1)}_{t|d}\right\|_2}{\left\|\boldsymbol{a}^{(k-1)}_{t|d}\right\|_2} < \epsilon$
```

provides four fast approximate filters which I will illustrate how to use with a hard disk failure data set.

	EKF (single iteration)	UKF	SMA	GMA
Approximation in correction step	Taylor	UT	Mode	Mode
Parallel	Yes	No	No	Yes
Depends on ordering	No	No	Yes	No
Additional hyperparameters	No	Yes	No	No
Sensitive to $\mathbf{Q}_0$	No	Yes	No	Yes

Table 1.1: Properties for the filter methods for line 4 of Algorithm 1. UT stands for unscented transform. The parallel row indicates whether the current implementation supports parallel computation. The “Depends on ordering” row indicates whether the method is sensitive to the ordering of the data set. The ”additional hyperparameters” indicates whether there are additional important hyperparameters with the method. The final row indicates whether the method often perform poorly if $\mathbf{Q}_0$ has large entries in the diagonal elements.

Table 1.1 shows the pros and cons of the methods. We make a Taylor expansion in the EKF given the $\boldsymbol{a}_{t|t-1}$ estimate from the prediction step. The UKF uses the so-called unscented transformation instead. This may yield a better approximation than the Taylor expansion. The SMA approximates the mode of $\boldsymbol{\alpha}_t$ given $\boldsymbol{a}_{t|t-1}$ and $\mathbf{V}_{t|t-1}$ adding the information of each observed outcome, y_{it}, in terms. The GMA does the same but uses all the observed outcomes, $\boldsymbol{y}_t$, at the same time. The UKF is currently not supporting parallel computation but could potentially. The simulation examples in Section 1.4 suggest that the EKF with multiple iterations and the GMA may be preferable. The methods will be covered in more detail in the following sections including examples of how to use with the **dynamichazard** package.

Data set example

I will use time until failure for hard disks as an example throughout this paper. Predicting when a hard disk will fail is important for any firm that manages large amounts of data stored locally to replace the hard disks before they fail. Self-monitoring, analysis, and reporting technology (SMART) is one tool used to predict future hard disk failures. The data set I will use is publicly available from BackBlaze (2017), which is a data storage provider that currently manages more than 65000 hard disks. Backblaze has a daily snapshot of the SMART attributes for all its hard disks going back to April 2013. The final data set is included with the package and has the name `"hds"`. Some minor changes[1] are made in this paper to the `"hds"` data set. The final data set I use has 79668 unique hard disks. It has 522041 rows in start-stop format for survival analysis.

A hard disk is marked as a failure if "... the drive will not spin up or connect to the OS, the drive will not sync, or stay synced, in a RAID Array ... [or] the Smart Stats we [Backblaze] use show values above our [Backblaze's] thresholds" (Klein, 2016). A hard drive with a failure is removed. I will not use the SMART attributes that Backblaze uses as covariates because of the third condition. These are SMART attributes 5, 187, 188, 197, and 198 (BackBlaze, 2014).

I will use the power-on hours (SMART attribute number 9) as the time variable in the model I estimate. The hard disks run 24 hours a day unless they are shut down (e.g., for maintenance). Thus, the power on hours reflects both the usage and age of the hard disk. The SMART attribute I will use as a predictor is the power cycle count (SMART attribute number 12). This counts the number of times a hard disk has undergone a full hard disk power on/off cycle. The power cycle count may be a proxy of batch effects as hard disks are stored in storage pods with 45 or 60 hard disks. An entire storage pod has to be shut down if a hard disk has to be replaced. Thus, the power cycle count may be a proxy for batch effects if hard disks from the same batch are stored in the same storage pods.

I will include a factor level for the hard disk version,[2] as the differences in failure rates between hard disk versions are large. In particular, one 3 terabyte (TB) Seagate hard disk version (ST3000DM001) has a high failure rate (Klein, 2015). I remove the 3 TB Seagate hard disks as only half of the hard disk that fails have a failure indicator set to one. I remove versions with

---

[1] I use last observation carried forward for the covariate, change the time scale to months, and I set time zero to 4 days of running.

[2] I write hard disk version instead of model to avoid confusion between a fitted statistical model and a hard disk model.

	$t \in (0, 20]$		$t \in (20, 40]$		$t \in (40, 60]$	
Hard disk version	#D	#F	#D	#F	#D	#F
ST4000DM000	36131	1036	12081	472	31	1
HMS5C4040BLE640	8511	34	3091	2	0	
HMS5C4040ALE640	7155	78	7077	11	44	
ST8000DM002	3927	13	0		0	
HDS5C3030ALA630	3864	19	4625	47	4532	52
HDS5C4040ALE630	2717	40	2665	33	2364	3
ST6000DX000	1915	35	45	1	0	
WD30EFRX	1284	126	876	22	129	1
ST500LM012 HN	800	24	147	2	0	
HDS723030ALA640	792	6	1040	30	997	23
WD60EFRX	495	36	253	12	0	
WD30EZRX	483	6	370	10	0	
ST31500541AS	150	14	712	49	1986	235
HDS722020ALA330	134	7	4765	46	4658	138
ST31500341AS	114	16	345	23	669	114
WD10EADS	38	2	124	8	463	16

Table 1.2: Summary information for each of the hard disk versions. The hard disk version is indicated by the first column. The number of disks is abbreviated as '#D' and total failures is abbreviated as '#F'. The $t \in (x, y]$ indicates which time interval the figures apply to. Blank cells indicate zeros.

fewer than 400 unique hard disks. These have either few cases or few observations. I winsorize at the 0.995 quantile for the power cycle count (i.e., I set values above the 0.995 quantile to the 0.995 quantile).

Table 1.2 provides information about each of the hard disk versions. The table shows that data is available only for some versions during parts of the 60-month period. Thus, some of the curves shown later will partly be extrapolation.

1.2.1 Extended Kalman Filter

The EKF approximates nonlinear state space models by making a given order Taylor expansion about the state vector, most commonly using the first order Taylor expansion. One of the EKF's advantages is that it results in formulas similar to the Kalman filter. The implemented algorithm is due to Fahrmeir (1992, 1994). The largest computational cost is the Taylor approximation which is $\mathcal{O}\left(q^2 n_t + q^3\right)$ where $n_t = |R_t|$ denotes the number of individuals at risk at time t. However, the computation is "embarrassingly parallel" because the computation can easily be separated into independent tasks that can be executed in parallel. This is exploited in the current version of `ddhazard` using the **thread** library in C++. The C++ **thread** library is a portable and standardised multithreading library.

Fahrmeir (1992) notes that the EKF is similar to a Newton-Raphson step, which can motivate us to take further steps. This is supported with the `ddhazard` function. Moreover, the EKF may have divergence problems. A learning rate (or step length) can be used in cases where divergence is a problem. Further, `ddhazard` increases the variance in the denominator of the terms used in the algorithm (see the "*ddhazard*" vignette in the package). This reduces the effect of values predicted near the boundaries of the outcome space. This is similar to the approach taken in **glmnet** (Friedman et al., 2010) to deal with large absolute values of the linear predictor.

One option is to take extra correction steps (extra a Newton-Raphson steps) which is not done by default. They will be taken if the `NR_eps` argument to `ddhazard_control` is set to the value of convergence threshold in the Newton-Raphson method. The user sets the learning rate (or step length) with the `LR` argument to `ddhazard_control`. By default, the current implementation tries a decreasing series of learning rates, starting with `LR` value until the algorithm does not diverge. The extra term in the denominators, as in the **glmnet** package, are set with the `denom_term` argument to `ddhazard_control`. Values in the range $[10^{-6}, 10^{-4}]$ tend to be sufficient in most cases. My experience is that the user should focus on the learning rate. See the "*ddhazard*" vignette in the package for the algorithm and further details.

I fit the model using the EKF with a single iteration in the correction step. I use a natural cubic spline for the number of power cycles to capture potential nonlinear effects. The code is shown below. First, I assign the `formula` using the `ns` function for a natural cubic spline:

```
R> library("splines")
R> library("dynamichazard")
R> frm <- Surv(tstart, tstop, fails) ~ -1 + model + ns(smart_12, knots = seq(3,
+    53, 10), Boundary.knots = c(0, 115))
```

I remove the intercept to not get a reference level for the disk version with the `-1`. Then I fit the model:

```
R> system.time(ddfit <- ddhazard(formula = frm, data = hd_dat, by = 1, max_T =
+    60, id = hd_dat$serial_number, Q_0 = diag(1, 23), Q = diag(.1, 23),
+    control = ddhazard_control(method = "EKF", eps = .001)))
```

```
   user  system elapsed
  54.84    5.55   15.86
```

`system.time` is used to show the computation time in seconds. `Q` is the initial value of the covariance matrix $\mathbf{Q}$ and `Q_0` is the covariance matrix $\mathbf{Q}_0$. The `by` argument in the call is used to specify the length of each interval. Thus, `by = 0.5` would give twice as many intervals, each with half the length. The last period we observe when estimating ends at `max_T`. `method = "EKF"` specifies that we want the EKF and `eps` is the convergence threshold used in the EM-algorithm.

I will focus on the first nine predicted parameters for the hard disk versions in this paper. Figure 1.1 shows the predicted parameters of the versions' factor levels. The plot shows the conditional log odds of failing in month t given survival up to time $t-1$ for a hard disk with zero power cycles. It is interesting that some versions seem to have a decreasing parameter for the factor in Figure 1.1, whereas the parameter increases for others. This can partly be explained by the "bathtub curve" used in reliability engineering. The bathtub curve is a hypothetical hazard curve which has a decreasing hazard rate to start with which is due to defective disks while later having an increasing hazard curve due to wear out. The early failures due to defective disks or later wear out may not be a factor for a particular version which can explain the curves

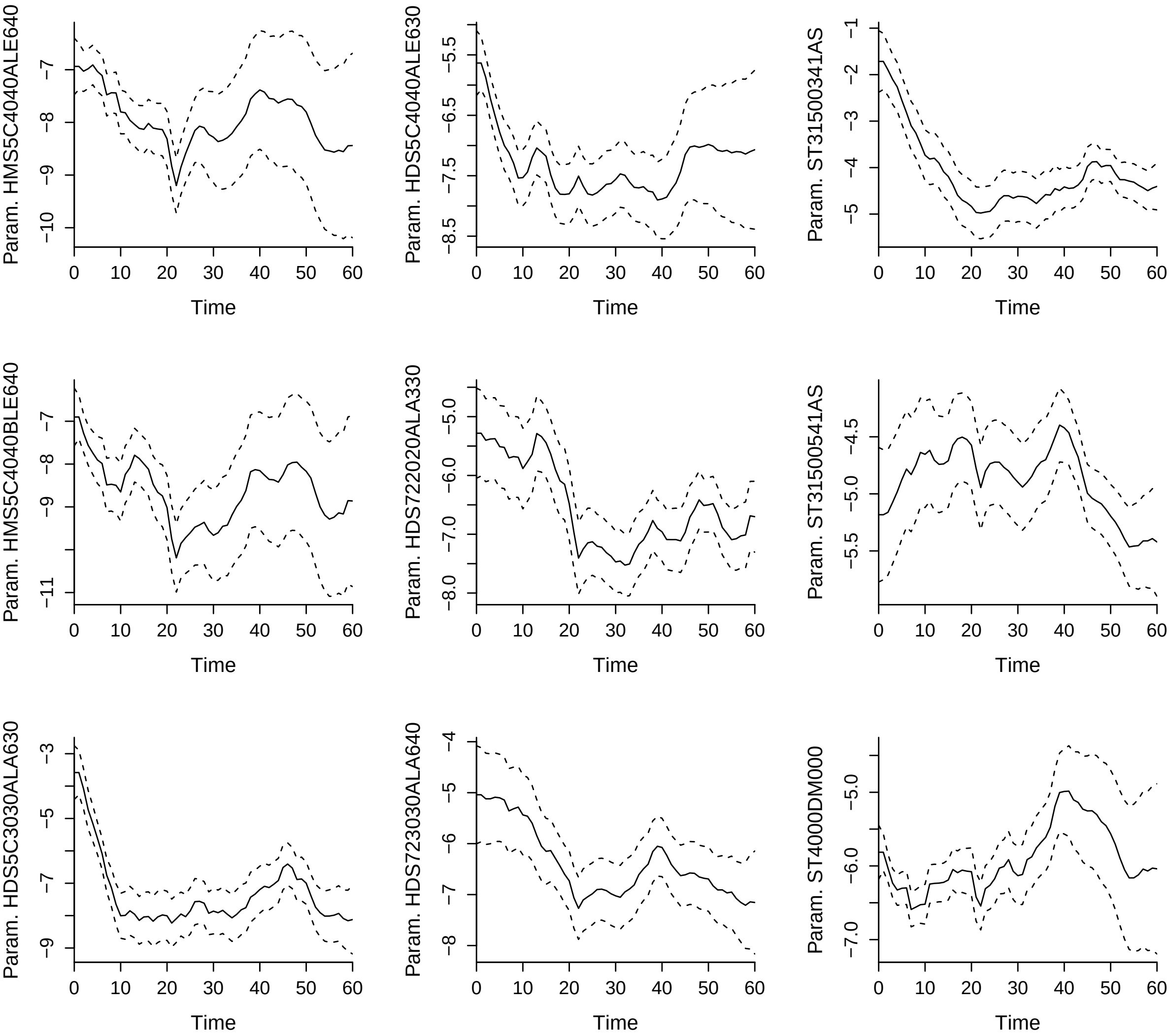


Figure 1.1: Predicted parameters for factor levels for the hard disk version with EKF with a single iteration in the correction step.

we see. Notice that some of the prediction intervals get wider or shorter in the start or at the end because of extrapolation. I only have data for at most three years for each hard disk. Furthermore, I only have data for some versions in parts of the 60-month period because of Backblaze's purchasing patterns. Thus, we see increasing or decreasing width of the prediction intervals for some parameters of factor levels in certain periods.

Figure 1.2 shows how the effect of the number of power cycles evolves over time for three specific choices of the power cycle count. It may seem odd that we do not have a monotone effect as we may expect the batch effect mentioned previously in which case we should see a monotonically increasing effect. However, some of the curves are not based on much data in some parts of the plot. E.g., it is not likely that a hard disk has had many power cycles in the start of the first few months.

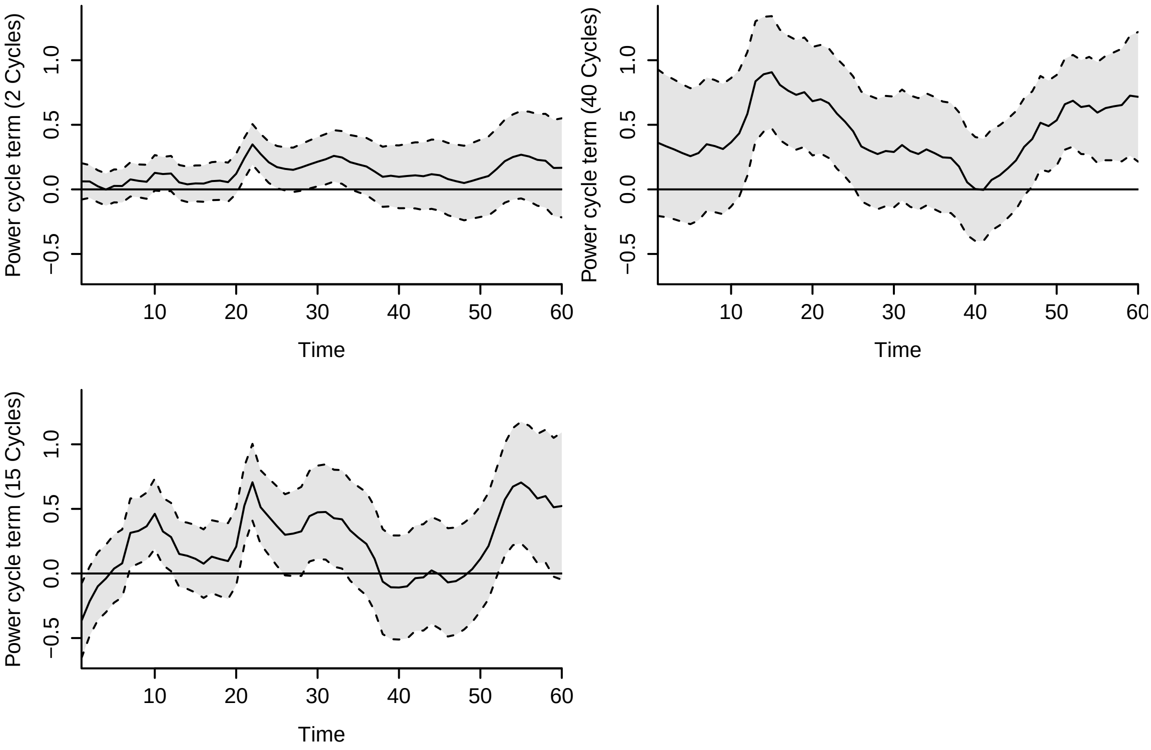


Figure 1.2: Plots of predicted terms on the linear predictor scale for different values of power cycle counts.

Examples with more iterations with the EKF

Next, I fit a model with more iterations in the correction step:

```
R> system.time(
+    ddfit_xtr <- ddhazard(formula = frm, data = hd_dat, by = 1, max_T = 60,
+    id = hd_dat$serial_number, Q_0 = diag(1, 23), Q = diag(.1, 23),
+    control = ddhazard_control(method = "EKF", eps = .001, NR_eps = .00001)))
```

```
   user  system elapsed
  189.1    21.3    44.6
```

`NR_eps` is the tolerance for the extra iterations in correction step. The default is `NULL` which yields only a single iteration. It takes longer due the additional correction steps. I plot the first nine factor levels with the following call:

```
R> for(i in 1:9){
+    plot(ddfit, cov_index = i)
+    plot(ddfit_xtr, cov_index = i, add = TRUE, col = "darkblue")
+    add_hist(i)
+  }
```

Figure 1.3 shows the plots. The `add_hist` is a function for the specific data set that adds the bars which heights reflect the relative number of individuals at risk in each interval for a given hard

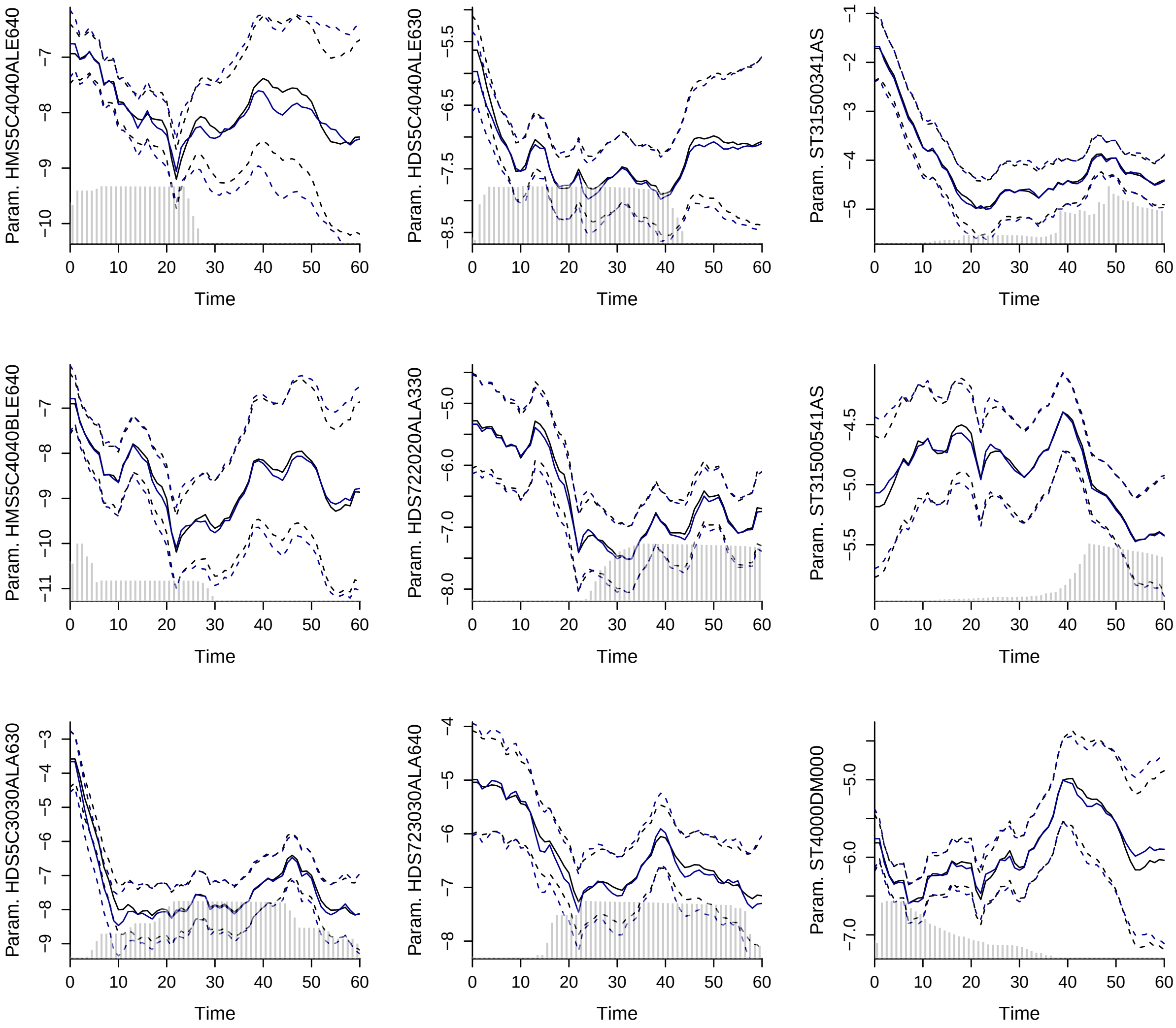


Figure 1.3: Predicted parameters with and without extra iterations in the correction step with the EKF. The blue curves are the estimate with extra iterations. Grey transparent bars indicate the number of individuals at risk for the specific hard disk version. Heights are only comparable within each frame.

disk version. For some of the parameters, the two plots differ noticeably, particularly the factors levels with a sparse amount of data (observations and/or failures) in some periods (cf., Table 1.2). A disadvantage of the EKF is that it may provide a poor approximation of the nonlinearities; This is what motivates the unscented Kalman filter in the next section.

1.2.2 Unscented Kalman Filter

The UKF is introduced by Julier and Uhlmann (1997). The idea is to select a fixed set of vectors and weights such that the mean and covariance matrix match those of the filtered state vector distribution at the prediction step. The points are then transformed in the correction step and an approximation of the mean and covariance matrix of the filtered state vector distribution of the next time step is then easily computed. The vectors and weights with the UKF are respectively known as *sigma points* and *sigma weights*. The UKF potentially provides a better approximation of the nonlinear dynamics than does the linear approximation used in the EKF. Further, the UKF does not require computation of the Jacobian. The latter advantage is not as important since deriving and computing the Jacobian is not complicated for the models in this paper.

The unscented transform performs the correction step by evaluating the conditional mean and covariance matrix of $\boldsymbol{y}_t$ at $2q+1$ weighted points, the so-called *sigma points*, given by:

$$\begin{aligned}
\widehat{\boldsymbol{a}}_0 &= \boldsymbol{a}_{t|t-1} \\
\widehat{\boldsymbol{a}}_j &= \boldsymbol{a}_{t|t-1} + \sqrt{q+\lambda}\left(\mathbf{V}_{t|t-1}^{1/2}\right)_j, \qquad j = 1,2,\dots,q \\
\widehat{\boldsymbol{a}}_{j+q} &= \boldsymbol{a}_{t|t-1} - \sqrt{q+\lambda}\left(\mathbf{V}_{t|t-1}^{1/2}\right)_j
\end{aligned} \tag{1.23}$$

with associated *sigma weights*:

$$W_0^{[m]} = \frac{\lambda}{q+\lambda} \tag{1.24}$$

$$W_0^{[c]} = \frac{\lambda}{q+\lambda} + 1 - \alpha^2 \tag{1.25}$$

$$W_j^{[m]} = W_j^{[c]} = \frac{1}{2(q+\lambda)}, \qquad j = 1,\dots,2q \tag{1.26}$$

where $\lambda = \alpha^2(q+\kappa) - q$, κ and α are hyperparameters, $W_j^{[m]}$ are weights used to compute the conditional mean of $\boldsymbol{y}_t$, and $W_0^{[c]}$ are weights used to compute the conditional covariance matrix of $\boldsymbol{y}_t$. $\mathbf{V}_{t|t-1}^{1/2}$ denotes the "square root" matrix of $\mathbf{V}_{t|t-1}$ and $\left(\mathbf{V}_{t|t-1}^{1/2}\right)_j$ denotes the jth column of the "square root" matrix. We can then evaluate an approximate conditional mean and covariance matrix of $\boldsymbol{y}_t$ by

$$\mathrm{E}\left(\boldsymbol{y}_t \mid \boldsymbol{y}_1,\dots,\boldsymbol{y}_{t-1}\right) \approx \overline{\boldsymbol{y}} = \sum_{j=0}^{2q} W_j^{[m]} \boldsymbol{z}_t\left(\widehat{\boldsymbol{a}}_j\right) \tag{1.27}$$

$$\mathrm{Var}\left(\boldsymbol{y}_t \mid \boldsymbol{y}_1,\dots,\boldsymbol{y}_{t-1}\right) \approx \sum_{j=0}^{2q} W_j^{[c]} (\widehat{\boldsymbol{y}}_j - \overline{\boldsymbol{y}})(\widehat{\boldsymbol{y}}_j - \overline{\boldsymbol{y}})^\top + \sum_{j=0}^{2q} W_j^{[c]} \mathbf{H}_t(\widehat{\boldsymbol{a}}_j) \tag{1.28}$$

`ddhazard` uses the Cholesky decomposition for the $\mathbf{V}_{t|t-1}^{1/2}$ decomposition. The hyperparameters on which the sigma points and sigma weights depend on can have values $0 < \alpha \leq 1$, $\kappa \in \mathbb{R}$ under the restriction that $q + \lambda = \alpha^2(q+\kappa) > 0$. The following example will provide an idea of the effect of the hyperparameters. Suppose that at time t in the correction step of the filter with a two dimensional state equation, $q = 2$, we have

$$\boldsymbol{a}_{t|t-1} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \qquad \mathbf{V}_{t|t-1} = \begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix}, \qquad \mathbf{V}_{t|t-1}^{1/2} = \begin{pmatrix} 1.41 & 0.707 \\ 0 & 0.707 \end{pmatrix} \tag{1.29}$$

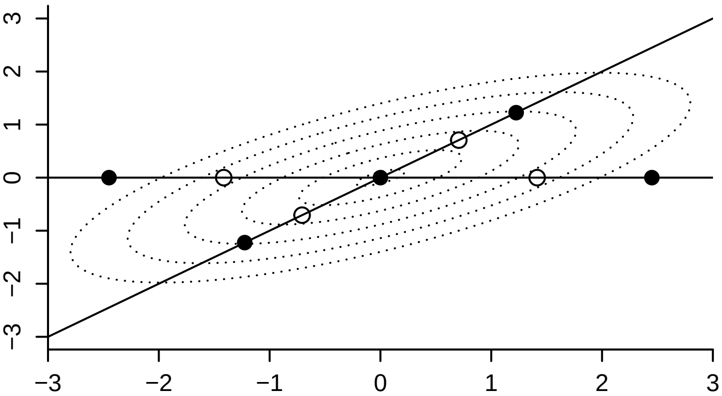


Figure 1.4: Illustration of sigma points in the example from Equation (1.29). The dashed lines are the contours of the density given by $\boldsymbol{a}_{t|t-1}$ and $\mathbf{V}_{t|t-1}$. The full lines are the direction given by the columns of the Cholesky decomposition. The filled circles are sigma points with $(\alpha, \kappa) = (1, 1)$ and the open circles are the sigma points with $(\alpha, \kappa) = (1/\sqrt{3}, 1)$. The point at $(0, 0)$ is a sigma point for both sets for hyperparameters.

Then the following hyperparameters yield to the following weights

$$\begin{aligned} (\alpha, \kappa) = (1, 1) \quad &\Rightarrow \quad \left(W_0^{[m]}, W_1^{[m]}, \dots, W_{2q}^{[m]}\right) = (1/3, 1/6, \dots, 1/6) \\ (\alpha, \kappa) = (1/\sqrt{3}, 1) \quad &\Rightarrow \quad \left(W_0^{[m]}, W_1^{[m]}, \dots, W_{2q}^{[m]}\right) = (-1, 1/2, \dots, 1/2) \end{aligned} \tag{1.30}$$

Decreasing α increases the absolute size of the weights of the last $2q$ sigma points and can lead to a negative weight on the zero sigma point, $\widehat{\boldsymbol{a}}_0$, as it does here. α also controls the spread of the sigma points through a $\alpha\sqrt{q+\kappa}$ factor in Equation (1.23). Decreasing α decreases the spread of the sigma points, as Figure 1.4 illustrates. The filled circles are the sigma points with $(\alpha, \kappa) = (1, 1)$, and the open circles are the sigma points with $(\alpha, \kappa) = (1/\sqrt{3}, 1)$.

A negative weight on the zero-th sigma point, $W_0^{[m]} < 0$, can cause computational issues, as Menegaz (2016) points out, since the conditional covariance matrix in Equation (1.28) can fail to be positive definite. Thus, we may select a specific value of $W_0^{[m]} > 0$ by setting $\kappa = q(1 + \alpha^2(W_0^{[m]} - 1))/(\alpha^2(1 - W_0^{[m]}))$ for a given value of α. The UKF can be tuned more than the EKF to any given data set while it may be hard to make estimation in an automatic fashion with the UKF.

`ddhazard` uses the three hyperparameter UKF given by Wan and Merwe (2000). There is an additional parameter denoted by β which is not included here for the sake of brevity. Many different UKFs have been suggested with different hyperparameters, algorithms, and sigma points (see Menegaz 2016 for a comparison of different forms of UKFs in the literature). Evaluating Equation (1.28) as in Wan and Merwe (2000) yields an $\mathcal{O}\left(n_t^3\right)$ computational complexity algorithm. It is reduced to $\mathcal{O}\left(n_t\right)$ with an application the Woodbury matrix identity. See the "*ddhazard*" vignette for further details.

Computation in parallel is not supported in the current version of `ddhazard` with the UKF. An identity matrix times a scalar is added to Equation (1.28) to reduce the effect of observation predicted near the boundary of the outcome space as done with the EKF. The scalar can be set with `denom_term` argument to `ddhazard_control`. I will end this section on the UKF with an example.

Examples with the UKF

One problem with the UKF compared with the EKF is its greater sensitivity to the choice of $\mathbf{Q}_0$ because it is used to form the sigma points at time zero. I will illustrate this in the following paragraphs. I fit the model below and plot the predicted parameters. I set $\mathbf{Q}_0$ to a diagonal matrix but with larger entries than before. I specify that I want the UKF by setting the argument `method = "UKF"` in the `ddhazard_control` call. `eps` is increased such that the methods is deemed to have converged within the default amount of maximum EM iterations.

```
R> system.time(ddfit_ukf <- ddhazard(formula = frm, data = hd_dat, by = 1,
+    max_T = 60, id = hd_dat$serial_number, Q_0 = diag(10, 23), Q = diag(.1,
+    23), control = ddhazard_control(method = "UKF", eps = .01)))
```

```
   user  system elapsed 
 59.704   0.384  59.334 
```

Figure 1.5 shows the result. Figure 1.6 shows the same model but with $\mathbf{Q}_0$'s diagonal entries equal to 0.1. The latter figure is comparable to what we have seen previously. A similar comment applies to the starting value of $\mathbf{Q}$. My experience is that we need to select a matrix that has *large* but not *too large* elements in the diagonal. See Xiong et al. (2006) for the covariance matrix role in a slightly different class of models. In contrast to the UKF, the EKF with one iteration in the correction step can have large entries in the diagonal of $\mathbf{Q}_0$.

1.2.3 Sequential Approximation of the Posterior Modes

Another idea is to replace the means in the filters with the modes in each correction step. That is, we are still looking for a method to perform the filtering in Algorithm 1. We perform the same prediction step as with the EKF and UKF, and we change the correction step from finding the mean to finding the mode. In making this replacement, we must find the minimum of Equation (1.31), followed by an update of the covariance matrix.

$$\boldsymbol{a}_{t|t} = \underset{\boldsymbol{\alpha}}{\arg\min} \left(-\log \mathrm{P}\left(\boldsymbol{\alpha} \;\middle|\; \boldsymbol{a}_{t|t-1}, \mathbf{V}_{t|t-1}\right) - \sum_{i \in R_t} \log \mathrm{P}\left(y_{it} \mid \boldsymbol{\alpha}\right) \right) \tag{1.31}$$

One way of finding an approximate minimum is to replace Equation (1.31) with $n_t = |R_t|$ rank-one updates of the form in Equation (1.32) and an update of the covariance matrix. I use a superscript to indicate the previous result from the rank-one update.

$$\boldsymbol{a}_{t|t}^{(k)} = \underset{\boldsymbol{\alpha}}{\arg\min} \left(-\log \mathrm{P}\left(\boldsymbol{\alpha} \;\middle|\; \boldsymbol{a}_{t|t}^{(k-1)}, \mathbf{V}_{t|t}^{(k-1)}\right) - \log \mathrm{P}\left(y_{i_k t} \mid \boldsymbol{\alpha}\right) \right) \tag{1.32}$$

I will refer to this method as the SMA. There are two implemented versions: one which makes the updates of the covariance matrix using the Woodbury matrix identity and one which updates

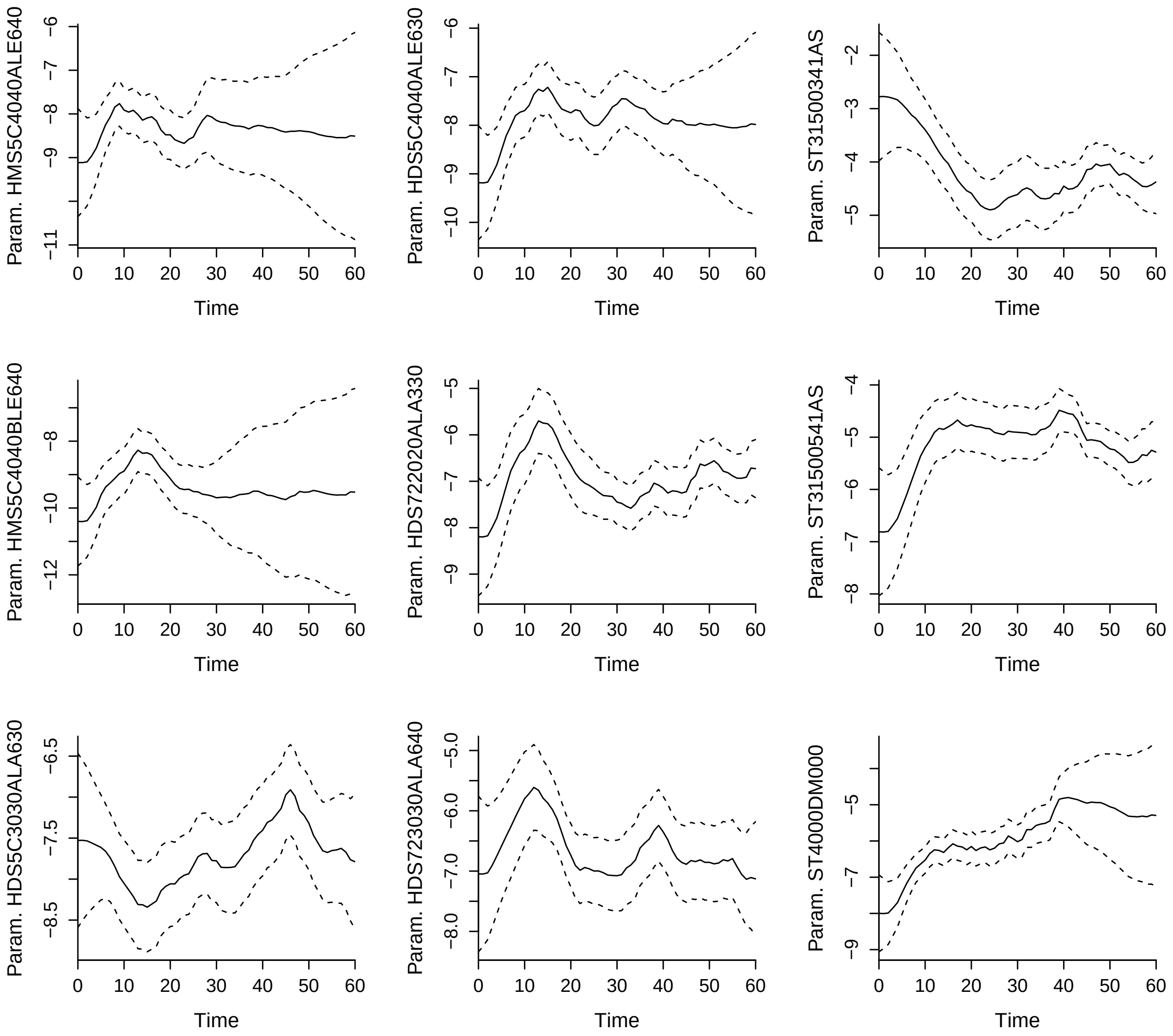


Figure 1.5: Predicted parameters with the UKF used on the hard disk failure dataset where $\mathbf{Q}_0$ has large entries in the diagonal.

a Cholesky decomposition of the concentration matrix instead. The latter guarantees that the covariance matrix is positive semi-definite but is slower. See the "*ddhazard*" vignette for further details.

The SMA can have large entries in the diagonal of $\mathbf{Q}_0$ like the EKF with one iteration. A disadvantage of SMA is that it is sequential and all matrix and vector products are in dimension $q \times q$ and q. Thus, although one could do the matrix operations in parallel then this is only advantageous if q is large. Moreover, the result depends on the order of the risk set. For this reason, the risk sets are permuted once before running the algorithm. This can be avoided by setting passing `permu = FALSE` in the `ddhazard_control` call.

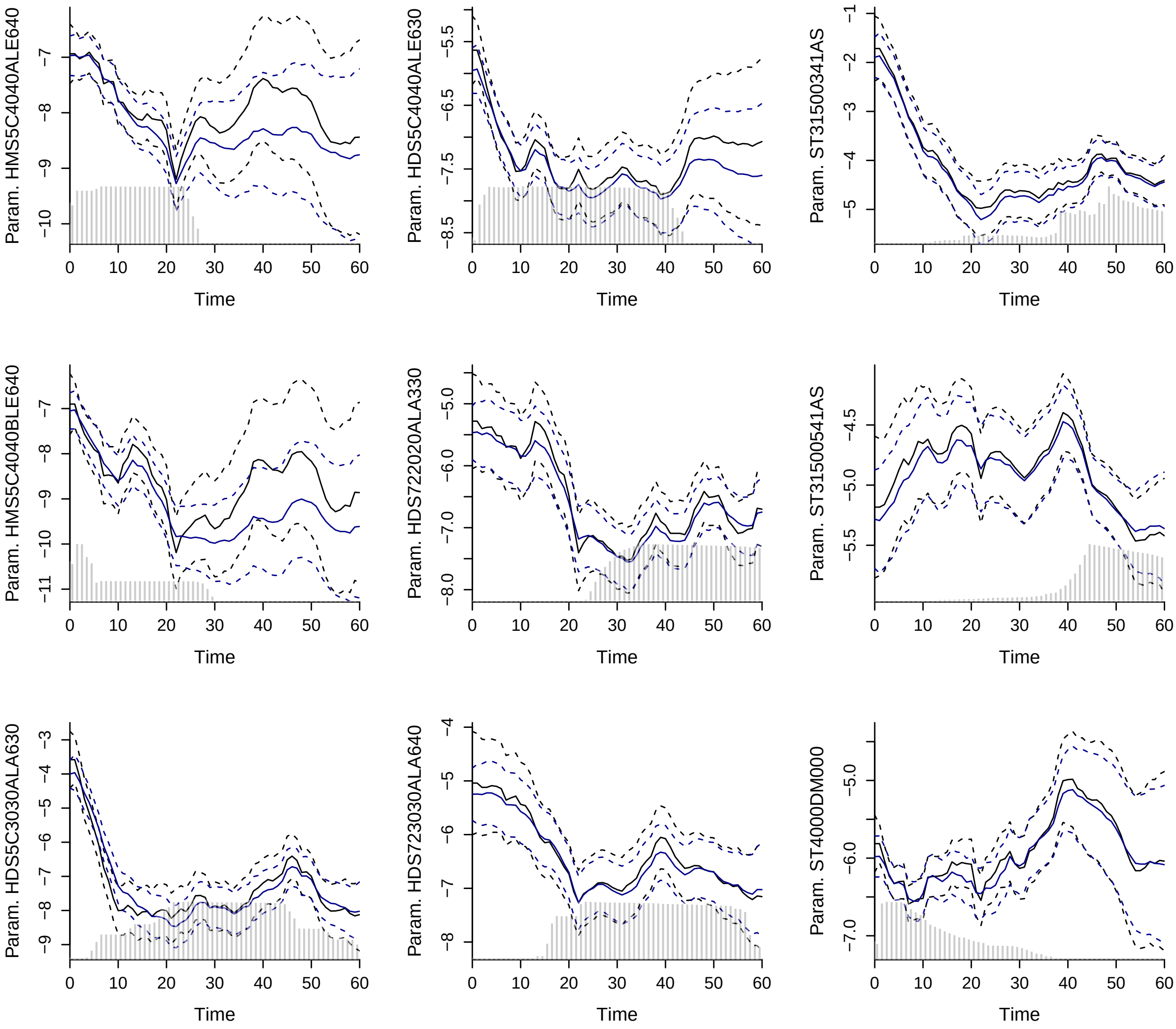


Figure 1.6: Similar plot to Figure 1.5 but where the diagonal entries of $\mathbf{Q}_0$ are 0.1. The black curve is the estimates from the EKF with one iteration in the correction step. Grey transparent bars indicate the number of individuals at risk for the specific hard disk version. Heights are only comparable within each frame.

1.2.4 Global Mode Approximation

We can also minimize the right-hand side of Equation (1.31) directly. This will be called the GMA method. It is equivalent to an L2 penalized generalized linear model (GLM) in every iteration. It can be shown that the EKF with more iterations solves an equivalent problem as a Newton-Raphson to minimize the right-hand side of Equation (1.31). Thus, details on the GMA is omitted here and can be found in the "*ddhazard*" vignette.

The GMA is sensitive to the choice of $\mathbf{Q}_0$ which works as an inverse penalty. To give an extreme example, suppose we have no events in the first interval and only an intercept. Setting $\mathbf{Q}_0$ to a diagonal matrix with large entries (in this case $\mathbf{Q}_0$ is a scalar) implies almost no restrictions on the intercept. Thus, selecting an intercept tending towards minus infinity is optimal. The computation with the GMA is done in parallel with **OpenMP** (Board, 2013).

The global mode approximation and the EKF with more iterations in the correction step are somewhat similar to the method in Durbin and Koopman (2012, Section 10.6). The major difference is that Durbin and Koopman (2012) make the Taylor expansion *before* running the filter about the current estimate of $\boldsymbol{\alpha}_0, \boldsymbol{\alpha}_1, \dots, \boldsymbol{\alpha}_d$ yielding so called pseudo-observations of an approximating Gaussian model. In contrast, the GMA method makes the expansion at each correction step *within* the filter about the current estimate of $\boldsymbol{a}_{t|t-1}$. The **KFAS** package implements the method in Durbin and Koopman (2012, Section 10.6). Further, **KFAS** uses the sequential method described in Koopman and Durbin (2000). The sequential method in **KFAS** is somewhat like the SMA. Again, the difference is at what point the Taylor expansion is made.

Examples with the SMA and GMA

I will use the hard disk failures data set to compare the SMA and GMA methods with the EKF with a single iteration in the correction step. Below, I estimate the model with the SMA method, and the GMA method. I use the correction step with the Cholesky decomposition with the SMA by setting the argument `posterior_version = "cholesky"` in the the `ddhazard_control` call.

```
R> system.time(ddfit_SMA <- ddhazard(formula = frm, data = hd_dat, by = 1,
+    max_T = 60, id = hd_dat$serial_number, Q_0 = diag(1, 23), Q = diag(.1,
+    23), control = ddhazard_control(eps = .001, method = "SMA",
+    posterior_version = "cholesky")))

   user  system elapsed
    443       0     442

R> system.time(ddfit_GMA <- ddhazard(formula = frm, data = hd_dat, by = 1,
+    max_T = 60, id = hd_dat$serial_number, Q_0 = diag(1, 23), Q = diag(.1,
+    23), control = ddhazard_control(eps = .001, method = "GMA")))

   user  system elapsed
  213.5     0.0    38.2
```

Figure 1.7 shows the three sets of predicted parameters. The parameters in Figure 1.7 appear similar. It is clear from the above that the SMA method using the Cholesky decompositions is much slower than the GMA. Further, the GMA has a computation time which is close to the EKF with more iterations shown earlier as expected.

1.2.5 Constant Effects

In some applications, constant (time-invariant) parameters may be relevant. A common way of estimating fixed parameters in filtering (e.g., see Harvey and Phillips, 1979) is to set the entries of the rows and columns of $\mathbf{Q}$ for the fixed parameters to zero and the corresponding diagonal entries of $\mathbf{Q}_0$ to large values. This approach is also used by Fahrmeir (1992) with the EKF. An alternative method to estimate the effects in the M-step is also included in the package but it is omitted here for brevity.

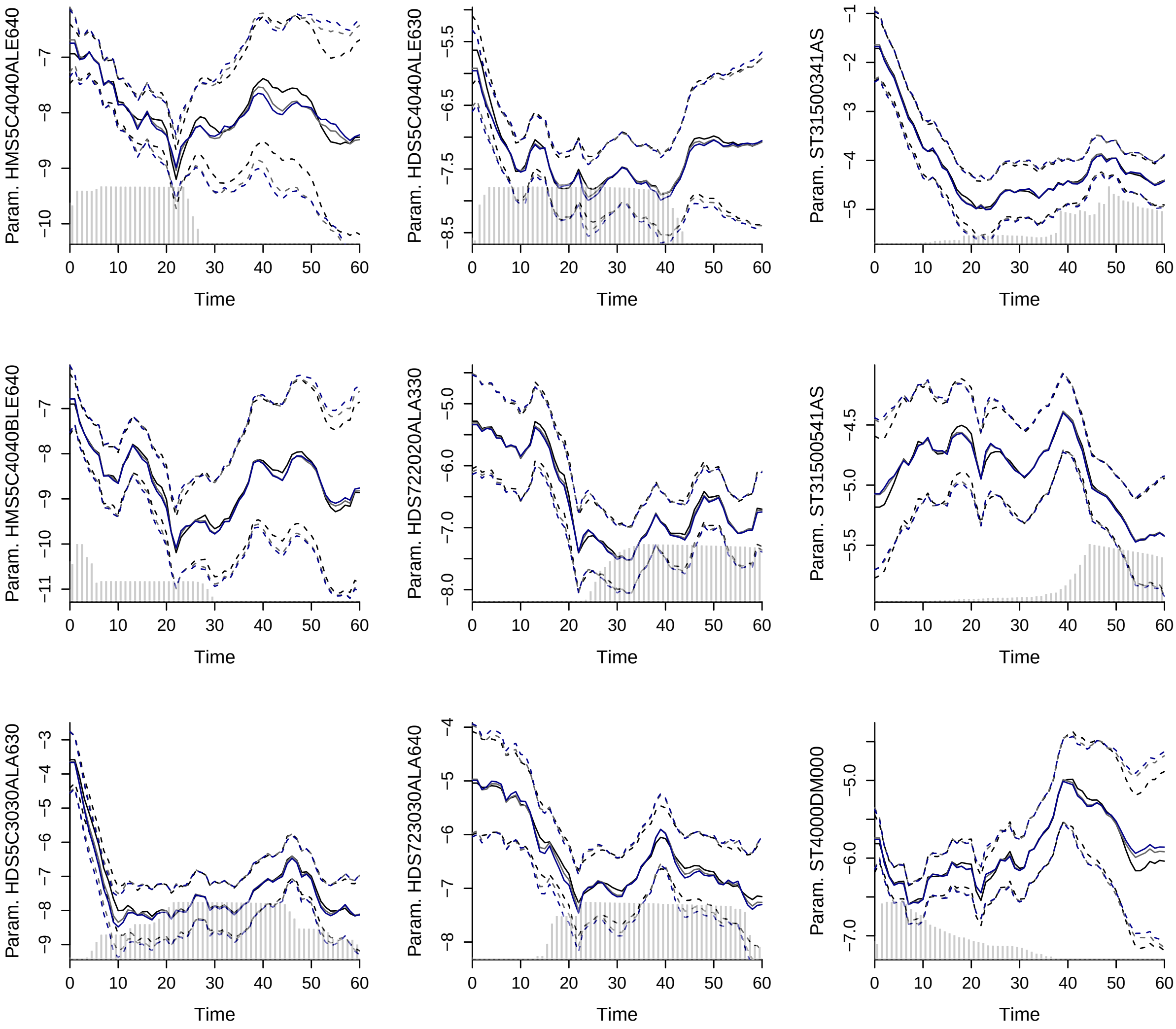


Figure 1.7: Predicted parameters using the EKF with a single iteration in the correction step, the GMA, and the SMA for the hard disk failure data set. The gray lines are the parameters from the SMA, blue lines are parameters from the GMA, and the black lines are the parameters from the EKF. Grey transparent bars indicate the number of individuals at risk for the specific hard disk version. Heights are only comparable within each frame.

1.2.6 Second Order Random Walk

I will end this part of the paper by estimating fixed parameters in the E-step as mentioned in the previous section. Further, I will illustrate the use of the second order random walk model. I estimate the model below where the factor levels for the hard disk version follow a second order random walk, and the spline for the SMART 12 attribute is fixed. I specify that I want a second order random walk for the factor levels by setting the argument `order = 2`. I specify which terms are fixed by wrapping the terms in the formula in the `ddFixed` function. The fixed effect estimation method is selected by setting `fixed_terms_method = "E_step"` in the `ddhazard_control` call. To avoid divergence, I decrease the learning rate by setting the `LR` argument the `ddhazard_control` call. Notice that `Q_0`'s dimension is twice that of `Q`. Lastly, I increase the maximum number of EM-iterations with the `n_max = 250` argument.

```
R> frm_fixed <- Surv(tstart, tstop, fails) ~ -1 + model + ddFixed(ns(smart_12,
+    knots = seq(3, 53, 10), Boundary.knots = c(0, 115)))
R> system.time(ddfit_fixed_E <- ddhazard(formula = frm_fixed, data = hd_dat,
+    by = 1, max_T = 60, order = 2, id = hd_dat$serial_number, Q_0 = diag(1,
+    32), Q = diag(.1, 16), control = ddhazard_control(method = "GMA", n_max =
+    250, NR_eps = .00001, eps = .001, LR = .1, fixed_terms_method =
+    "E_step")))
```

```
    user   system  elapsed
3626.355    0.614  609.564
```

Figure 1.8 shows the predicted factor levels for the hard disk version, and Figure 1.9 shows the spline estimate. The curves are more smooth compared to the first order random walk model as expected. The spline estimate shows a close to monotone increasing effect of the number of power cycle as we may have expected.

1.3 Discrete Versus Continuous Time

The dynamic discrete time model is where we use the log-likelihood terms, $l_{it}(\boldsymbol{\alpha}_t)$, as shown in Equation (1.21) where h is the inverse logit function, $h(x) = \exp(x)/(1+\exp(x))$. This model is suited for situations where the events are observed in intervals and the covariates change at discrete times. However, this is not the case for the hard disk data set. The hard disk data is not reported on monthly precision but on hourly precision. I print the first 10 rows here to illustrate this:

```
R> hd_dat[1:10, c("serial_number", "model", "tstart", "tstop", "smart_12")]
```

```
    serial_number         model tstart  tstop smart_12
505      5XW004AJ ST31500541AS 30.001 40.010        0
506      5XW004AJ ST31500541AS 40.010 43.172       24
507      5XW004AJ ST31500541AS 43.172 56.917       25
508      5XW004Q0 ST31500541AS 40.618 50.962        0
509      5XW004Q0 ST31500541AS 50.962 53.729       54
510      5XW004Q0 ST31500541AS 53.729 54.122       56
511      5XW004Q0 ST31500541AS 54.122 54.424       57
512      5XW004Q0 ST31500541AS 54.424 54.457       58
513      5XW004Q0 ST31500541AS 54.457 54.690       59
514      5XW004Q0 ST31500541AS 54.690 57.193       61
```

I will explain how the `ddhazard` implementation discretizes continuous event times in the following paragraphs. I redefine $\boldsymbol{x}_{ij}$ as the covariate vector for individual i in the period $(t_{i,j-1}, t_{ij}]$. Next, I redefine the discrete time risk set, R_t, as

$$R_t = \{(i,j) : \, t_{i,j-1} \leq t-1 < t_{i,j} \wedge D_{ij} = 0\} \tag{1.33}$$

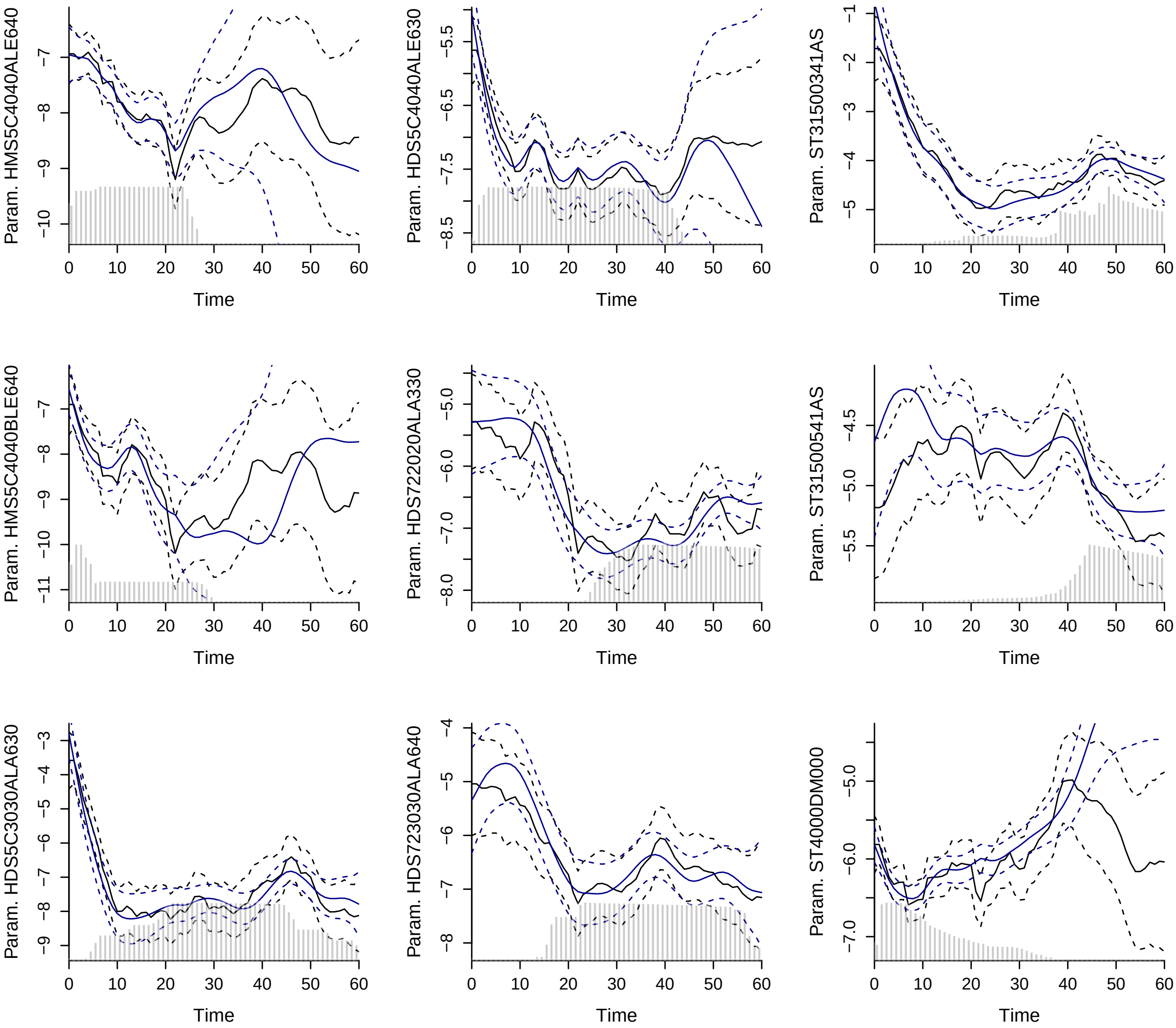


Figure 1.8: Predicted parameters for some of the factor levels with the second order random walk model. The blue lines are parameters with the second order random walk model with a fixed effect for the power cycle count and the black lines are parameters with the first order random walk model where all parameters are time-varying. Grey transparent bars indicate the number of individuals at risk for the specific hard disk version. Heights are only comparable within each frame.

Further, I redefine

$$y_{ijt} = 1_{\{T_i \in (\max(t_{i,j-1},t-1),\min(t_{ij},t)]\,\wedge\, t-1<t_{ij}\leq t\}} \tag{1.34}$$

$$l_{ijt}(\boldsymbol{\alpha}_t) = y_{ijt}\log h(\boldsymbol{x}_{ij}^\top\boldsymbol{\alpha}_t) + (1-y_{ijt})\log\Big(1-h(\boldsymbol{x}_{ij}^\top\boldsymbol{\alpha}_t)\Big) \tag{1.35}$$

y_{ijt} is a generalization of Equation (1.12) that indicates whether individual i experiences an event with *the jth covariate vector* in interval t. The following example will illustrate the impact of discrete time risk sets in Equation (1.33). Suppose we look at interval $d-1$ and d (the last two intervals) in a model with time-varying covariates. Further, let both the event times and the point at which we observe new covariates happen at continuous points in time.

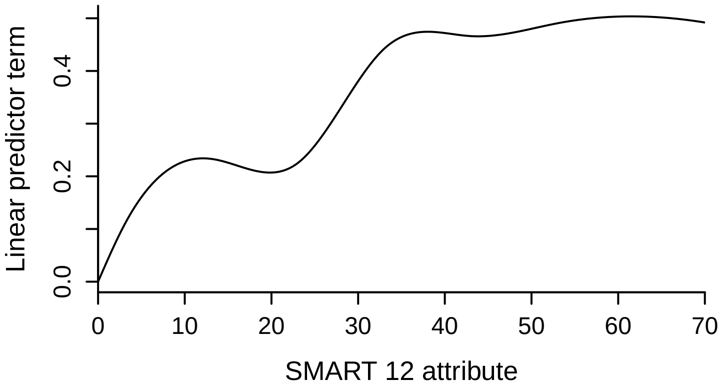


Figure 1.9: Fixed effects estimates for the SMART 12 attribute on the linear predictor scale.

Figure 1.10 illustrates such a situation. Each horizontal line represents an individual. A cross represents when the covariate values jump for the individual, and a filled circle represents an event that has happened for the individual. Lines that end with an open circle are right censored. The vertical dashed lines in the figure represent the time interval borders. The first vertical line from the left is the start of interval $d-1$, the second vertical line is when interval $d-1$ ends, and interval d starts, and the third vertical line is when interval d ends. I will use observation a in Figure 1.10 to illustrate the risk set in Equation (1.33). The covariate vector used in interval $d-1$ is $\boldsymbol{x}_{a1}$ as $t_{a0} < d-2 < t_{a1}$. By similar arguments, the covariate vector in interval d is $\boldsymbol{x}_{a2}$.

Because we use the risk set in Equation (1.33), we use covariates from 1 for individuals a, c, d, and f for the entire period of interval $d-1$, even though the covariates change at 2. Furthermore, g is not in either interval, as we only know that it survives parts of interval $d-1$. Lastly, we never include b as we do not know its covariate vector at the start of interval d.

1.3.1 Continuous Time Model

The continuous time model implemented in `ddhazard` is the model shown in Equation (1.11) in the introduction. The assumptions of the model are that

- the instantaneous hazards are given by $\exp(\boldsymbol{x}_i(t)^\top \boldsymbol{\alpha}(t))$.
- parameters jump at the end of time intervals, i.e., $\boldsymbol{\alpha}(t) = \boldsymbol{\alpha}_{\lceil t \rceil}$ where $\lceil t \rceil$ gives the ceiling of t. This is illustrated in Figure 1.10 where the parameters jump at the vertical lines.
- the individuals' covariates jump, i.e., $\boldsymbol{x}_i(t) = \boldsymbol{x}_{ij}$ where $j = \{k : \; t_{i,k-1} < t \leq t_{i,k}\}$. In Figure 1.10, the covariates jump at the crosses.

The instantaneous hazard jumps when either the individual's covariates jump or the parameters jump. Thus, an individual's event time is piecewise exponentially distributed given the state vectors. The log-likelihood of individual i having an event at time t_i is

$$\log\left(L\left(t_i \mid \boldsymbol{\alpha}_0, \dots, \boldsymbol{\alpha}_d\right)\right) = \boldsymbol{x}_i(t_i)^\top \boldsymbol{\alpha}(t_i) - \int_0^{t_i} \exp\left(\boldsymbol{x}_i(u)^\top \boldsymbol{\alpha}(u)\right) du \qquad (1.36)$$

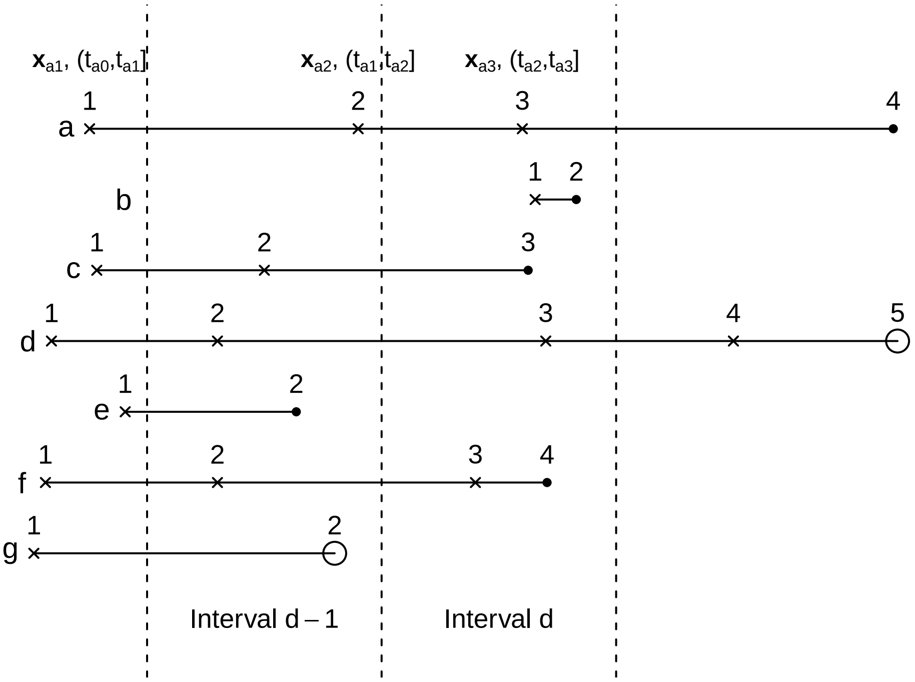


Figure 1.10: Illustration of a data set with 7 individuals with time-varying covariates. Each horizontal line represents an individual. Each number indicates a start time and stop time in the initial data. A cross indicates that new covariates are observed while a filled circle indicates that the individual has an event. An open circle indicates that the individual is right censored. Vertical dashed lines are time interval borders. The symbols for the covariate vectors and stop times are shown for observation a.

where $L(\cdot)$ denotes the likelihood. Because of our assumptions, the complete data log-likelihood in Equation (1.20) simplifies to

$$\begin{aligned}\mathcal{L}\left(\mathbf{Q},\mathbf{Q}_0,\boldsymbol{\mu}_0\right) = &-\frac{1}{2}\left(\boldsymbol{\alpha}_0-\boldsymbol{\mu}_0\right)^\top\mathbf{Q}_0^{-1}\left(\boldsymbol{\alpha}_0-\boldsymbol{\mu}_0\right)\\ &-\frac{1}{2}\sum_{t=1}^{d}\left(\boldsymbol{\alpha}_t-\mathbf{F}\boldsymbol{\alpha}_{t-1}\right)^\top\mathbf{R}\mathbf{Q}^{-1}\mathbf{R}^\top\left(\boldsymbol{\alpha}_t-\mathbf{F}\boldsymbol{\alpha}_{t-1}\right)\\ &-\frac{1}{2}\log\left|\mathbf{Q}_0\right|-\frac{d}{2}\log\left|\mathbf{Q}\right|\\ &+\sum_{t=1}^{d}\sum_{(i,j)\in\mathcal{R}_t}l_{ijt}(\boldsymbol{\alpha}_t)+\dots\end{aligned} \tag{1.37}$$

where

$$l_{ijt}(\boldsymbol{\alpha}_t) = y_{ijt}\boldsymbol{x}_{ij}^\top\boldsymbol{\alpha}_t-\exp\left(\boldsymbol{x}_{ij}^\top\boldsymbol{\alpha}_t\right)\left(\min\{t,t_{ij}\}-\max\{t-1,t_{i,j-1}\}\right) \tag{1.38}$$

The l_{ijt} terms are a simplification of Equation (1.36), where I use the assumption that both the covariates, $\boldsymbol{x}_i(t)$, and parameters, $\boldsymbol{\alpha}(t)$, are piecewise constant. Further, $\mathcal{R}_t$ is the *continuous time risk set* given by

$$\mathcal{R}_t = \left\{(i,j) : t_{i,j-1} < t \wedge t_{ij} \geq t-1 \wedge D_{ij} = 0\right\} \tag{1.39}$$

The two conditions in $\mathcal{R}_t$ are that the observation must start before the interval ends ($t_{i,j-1} < t$), and end after the interval starts ($t_{ij} \geq t-1$). I will use observation a in Figure 1.10 as an example. Observation a has two covariate vectors in interval $d-1$. The first is $\boldsymbol{x}_{a1}$ as $t_{a0} < d-1$ and $t_{a1} > d-2$. Similar arguments apply for the covariate vector $\boldsymbol{x}_{a2}$.

Example with the continuous model

As mentioned in Section 1.3, the start and stop times in the hard disk failure data set are in fractions of months on a hourly precision. Thus, I can use the continuous model. I fit the model with the EKF with more iterations in the correction step and get the continuous model by setting `model = "exponential"`:

```
R> system.time(ddfit_cont <- ddhazard(formula = frm, data = hd_dat, by = 1,
+    max_T = 60, model = "exponential", id = hd_dat$serial_number, Q_0 = diag(
+    1, 23), Q = diag(.1, 23), control = ddhazard_control(NR_eps = .0001, eps =
+    .001, LR = .5, method = "EKF")))
```

```
   user  system elapsed
  117.6    12.2    26.0
```

Figure 1.11 shows the first estimated factor levels' parameters. The results are comparable to what we have seen previously (e.g., see Figure 1.3 where I also used the EKF with more iterations in the correction step but with the discrete time model).

1.4 Simulations

In this section, I will simulate data using the first order random walk model and illustrate the computation time, and mean square error (MSE) of the predicted parameters as the number of individuals increases. I use a first order random walk for the parameters with 21 parameters. The intercept starts at -3.5, and the other parameters start at points drawn from the standard normal distribution. I set the intercept to a low value to decrease the baseline likelihood of an event in every interval. I let covariance matrix $\mathbf{Q}$ be a diagonal matrix which has 0.1^2 in the first entry (the intercept) and 0.33^2 for rest of the diagonal entries. The standard deviation is chosen lower for the intercept to ensure that the intercept does not change "too much" with high probability. Figure 1.12 provides an example of a draw of parameters.

I simulate a different number of individuals with $n = 2^{10}, 2^{11}, \dots, 2^{18}$ in each trial. Each individual is right censored at time 30, and I set the interval lengths to 1. Further, I simulate random delayed entry. We randomly start to observe each individual at time $0, 1, \dots, 29$ with a 50% chance of 0 and uniform chance on the other points. This mimics a situation like corporate default prediction where we use calendar time as the time scale. A firm may first be incorporated a while into the study, in which case the firm is subject to delayed entry.

Each individual has time-varying covariates that change after five periods. Thus, if an individual starts at time 2, his covariate vector changes at time $7, 12, \dots, 27$. The covariates are drawn from an iid standard normal distribution. For each value of n, I make 11 simulation trials.

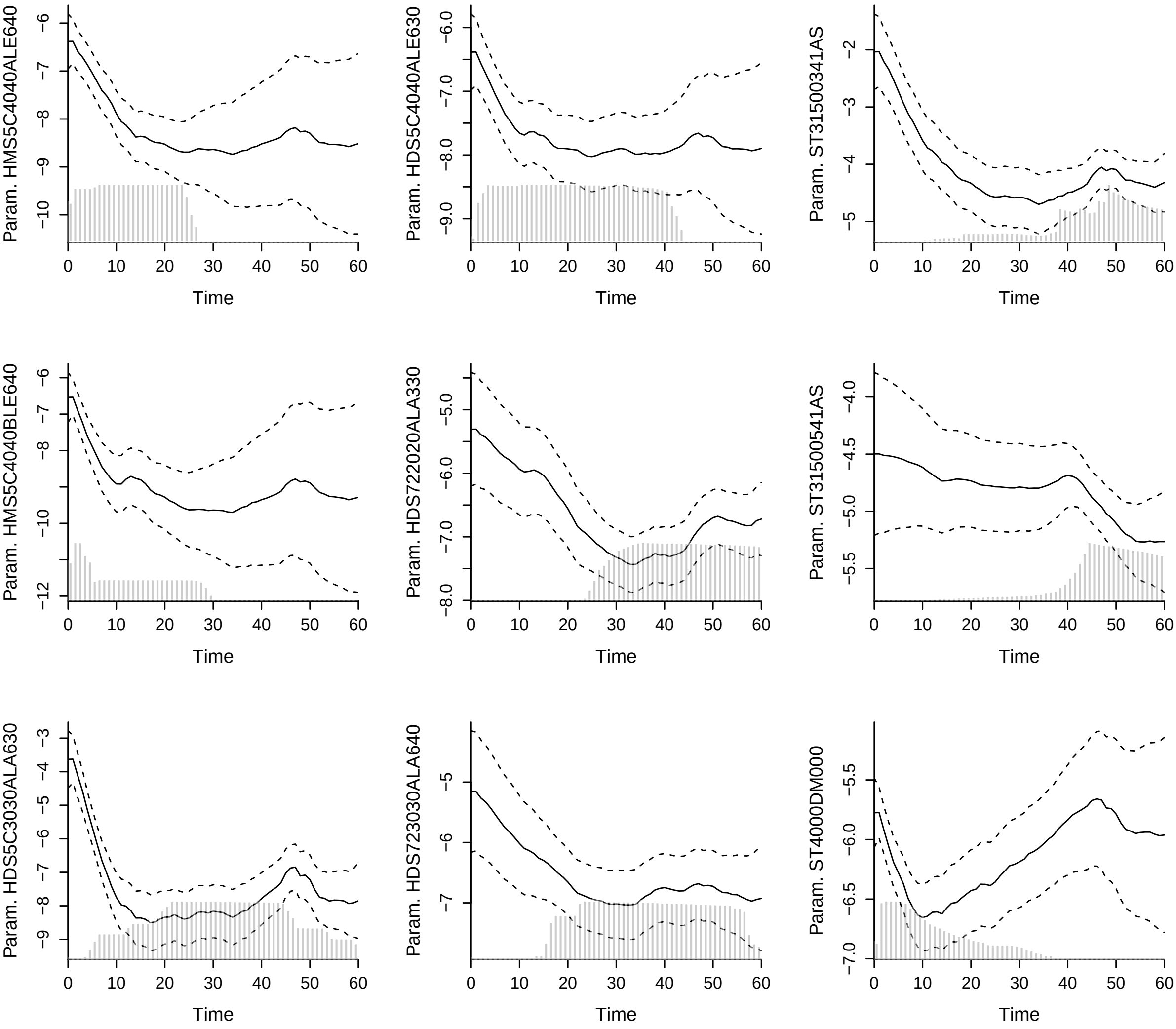


Figure 1.11: Predicted factor levels parameters with the continuous time model.

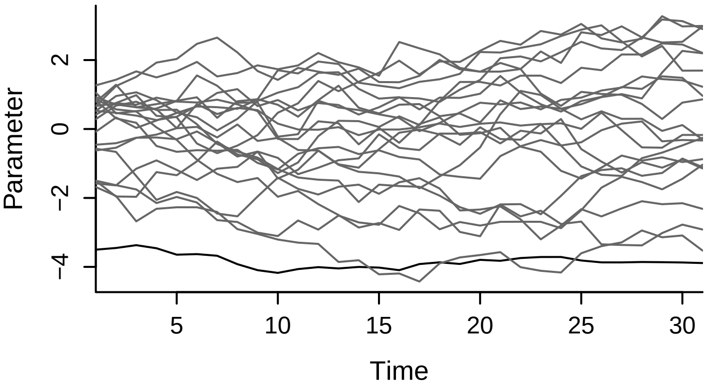


Figure 1.12: Example of parameters in the simulation experiment. The black curve is the intercept and the gray curves are the parameters for the covariates.

	EKF	EKF with extra iterations	UKF	SMA	GMA
Run time	2.324	4.763	15.839	10.550	5.196
Log-log slope	0.760	0.812	1.042	0.778	0.805

Table 1.3: Summary information of the computation time in the simulation study. The first row shows the median runtime for largest number of individuals. The UKF is only up to $n = 32768$. The second row shows the slope of the log computation time regressed on the log number of individuals for $n \geq 16384$.

I estimate the UKF model only up to $n = 2^{15}$ because of the computation time. Further, I set the UKF hyperparameters to $(\alpha, \beta, \kappa) = (1, 0, 0.004)$, which yields $W_0^{[m]} = 0.0001$. $\mathbf{Q}_0$ for the EKF with extra iterations, and the GMA is a diagonal matrix with entries 1. The UKF has 0.01 as the diagonal entries. The EKF without extra iterations and the SMA have 10000 in the diagonal entries of $\mathbf{Q}_0$. All the filters have the starting value of $\mathbf{Q}$ as a diagonal matrix with 0.01 in the diagonal elements. All the methods take at most 25 iterations of the EM-algorithm if the convergence criteria is not previously met.

The simulations are run on a laptop with Ubuntu 18.04 with an Intel® core™ i7-8750H @ 2.20GHz and 16GB ram. Figure 1.13 shows the medians and means of the computation time. Table 1.3 displays the median computation time for the largest value of n along with the regression slope of the log computation time regressed on the log number of individuals. All methods have a slope close to or below 1, reflecting the $\mathcal{O}(n_t)$ computational complexity. In fact, the slope is less than 1 for all but the UKF method. This can be explained by the overhead of the parallel computation. Further, the methods tend to use less EM iterations when more data is available. The latter can be seen from Figure 1.15, which shows the median number of iterations of the EM-algorithm. All the computation times include the time required to set up the model matrix and fit a weighted GLM to get a starting values for $\boldsymbol{\alpha}_0$. The setup time is equal for all methods.

Figure 1.14 shows a plot of the MSE for the parameters. The EKF with one iteration in the correction step does not improve much as n increases. Hence, more iterations seem preferable in this example. Some points are worth stressing. First, the computation time of the UKF can be reduced by using a multithreaded **BLAS** library or reimplementing the code. I have seen a reduction up to factor 2 for larger data sets on the setup used in the simulation when **OpenBLAS** (Xianyi et al., 2012) is used. Further, one can do more tuning (especially with the UKF) for each data set, which is not done in the present case.

The simulation here is "extreme" in that the linear predictor can take large absolute values in the last intervals with a nonnegligible probability. Thus, I perform a second simulation experiment where I draw the covariates from a normal distribution with zero mean and variance 0.33^2. Figure 1.16 shows MSEs. The difference between the filters is small in terms of mean square error. The computation times are similar to before in terms of the relative differences. Still, it seems that the EKF with extra iterations, and the GMA are preferred.

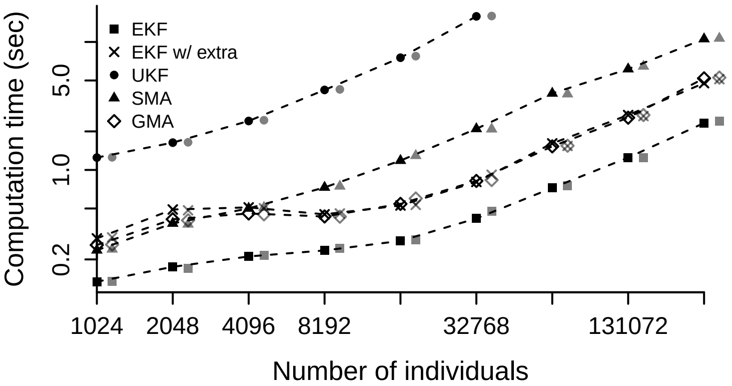


Figure 1.13: Median computation times of the simulations for each method for different values of n. The gray symbols to the right are the means. The filled squares are the EKF, the crosses are the EKF with extra iteration, the circles are the UKF, the triangles are the SMA, and the open squares are the GMA. The scales are logarithmic so a linear trend indicates that computation time is a power of n.

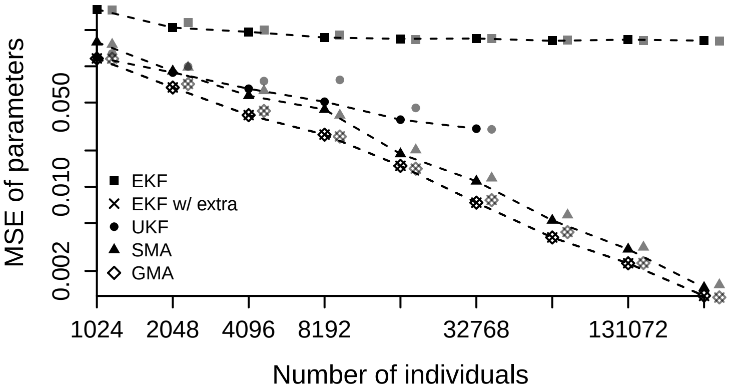


Figure 1.14: Median mean square error of predicted parameters of the simulations for each method for different values of n. The gray symbols to the right are the means. The filled squares are the EKF, the crosses are the EKF with extra iteration, the circles are the UKF, the triangles are the SMA, and the open squares are the GMA. The axis are on the logarithmic scale.

1.5 Comparison with Other Packages

I will end by comparing the implemented methods with other packages in R mentioned in the start of the article. Specifically, I will discretize the time line and use the `gam` function in the **mgcv** package to get a time-varying parameter by adding a spline over time for the intercept on the log odds scale as described in Tutz and Schmid (2016, chapter 5). Further, I use the `phreg` function from the **eha** package with argument `dist = "pch"` to get a piecewise constant intercept in an exponentially distributed arrival time model. We set each time interval to have length 1. I will also estimate a Cox model with the `coxph` function from the **Survival** package. Lastly, I use the discrete model with logit link function from `ddhazard` to get comparable results to the `gam` function. We use interval widths 0.5 for both `ddhazard` and `gam`.

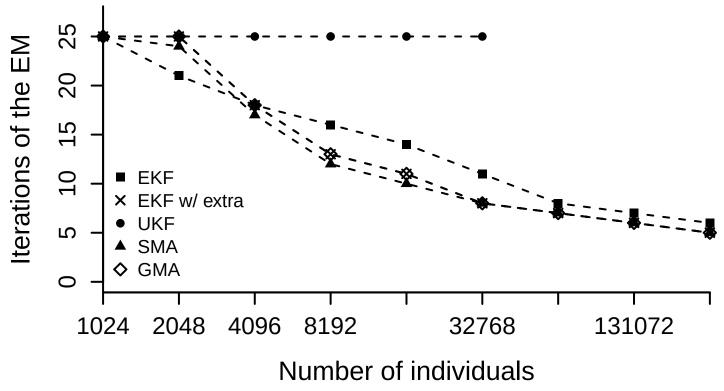


Figure 1.15: Median number of iterations of the EM-algorithm. The filled squares are the EKF, the crosses are the EKF with extra iteration, the circles are the UKF, the triangles are the SMA, and the open squares are the GMA.

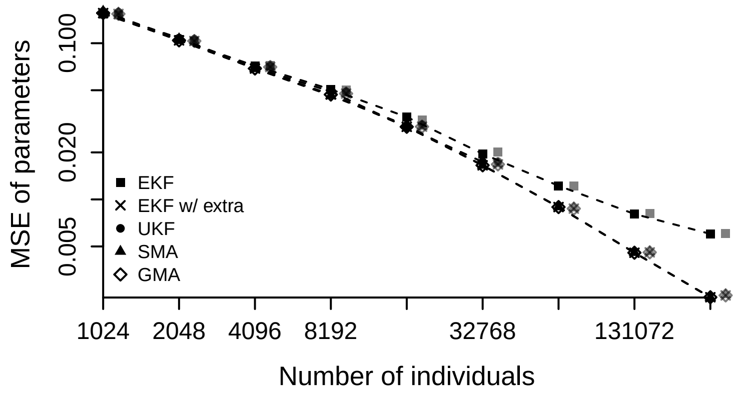


Figure 1.16: Similar plot to Figure 1.14 but where each element of the covariate vectors is drawn from $N\left(0, 0.33^2\right)$.

For simplicity, we focus on the ST4000DM000 hard disk version. Further, we take the hard disk information up to month 35. We do not have data for most of the hard disks beyond this time point. We only estimate a linear effect on the link scale in each model for the power cycle count with a time constant parameter.

The left plot of Figure 1.17 shows the discrete hazard rates with the power cycle count equal to zero. It shows that the result from `gam` and `ddhazard` are similar. The estimate from `phreg` is less smooth as there is a separate coefficient for each time interval of length 1 without any penalties. We use a natural cubic spline for the intercept in the `gam` fit. Thus, the log odds are linear in time beyond month 35 as these splines are linear beyond the boundary knots. This results in an almost linear increasing discrete hazard as the inverse logit function is almost linear for small inputs. Moreover, the prediction intervals are wider for the `gam` fit than for the `ddhazard` fit. The width of the prediction intervals with `ddhazard` fit mainly depends on the estimate of $\mathbf{Q}$ which is estimated to fit the whole curve.

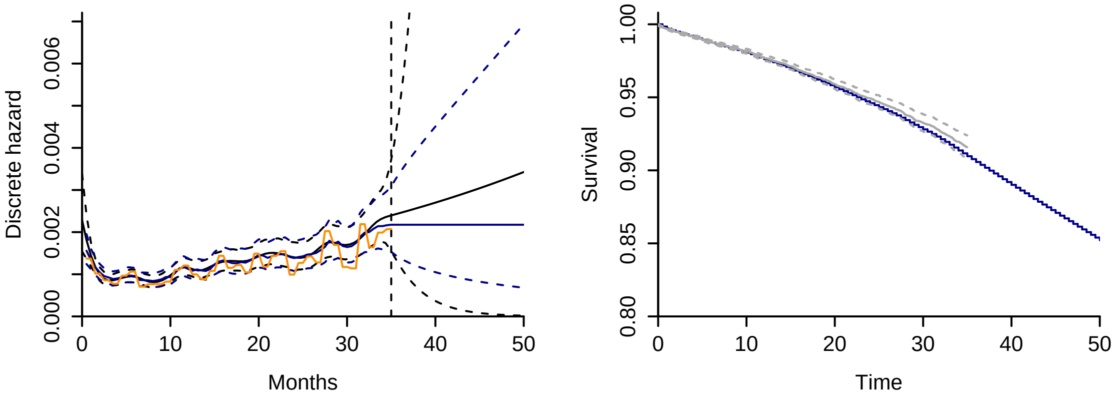


Figure 1.17: The left plot shows the discrete hazard rate when the power cycle count is equal to zero. The black line is the rate from `mgcv::gam`, the blue line is the rate from `dynamichazard::ddhazard`, and the orange line is the rate from `eha::phreg`. The dashed lines are 95% back-transformed prediction intervals (confidence bounds are not provided by `eha::phreg`). The right plot shows the estimated survival curve with the power cycle count equal to zero. The dark gray line is the curve from `survival::coxph` with 95% prediction intervals and the blue curve is from `dynamichazard::ddhazard`.

The right plot of Figure 1.17 shows the estimated survival curve from the Cox model from the `coxph` call and dynamic discrete hazard model from the `ddhazard` call both with the power cycle count equal to zero. The former is computed with the `survfit` function while the latter is computed with the `ddsurvcurve` function. The two generally agree but only the dynamic discrete hazard model allows for extrapolation beyond time 35 as the baseline hazard is nonparametric in the Cox model. The curve from `ddsurvcurve` is illustrated as a step function as the model provides discrete hazard rates.

1.6 Conclusion

I have covered the EM-algorithm used in the `ddhazard` function in **dynamichazard** and the four different filters available with the `ddhazard` function, highlighting the pros and cons of the different filters. Further, I have covered the dynamic discrete time model and the dynamic continuous time model. The simulation study shows that the filters scale well with the number of observations and are fast. Further, the simulation study shows how the mean square error of the predicted parameters behaves with different numbers of observations. The extended Kalman filter has been compared with other methods in R.

I have not covered all the `S3` methods provided in the **dynamichazard** package. These include `plot`, `predict`, `hatvalues`, and `residuals`. It is possible to include weights for the observations with all the filters. The details hereof are in the "*ddhazard*" vignette of this package. Furthermore, the `ddhazard_boot` function can be used to perform a nonparametric bootstrap. Weights are used in `ddhazard_boot` with case resampling, which reduces the computation time. Vignettes are provided with the **dynamichazard** package which illustrate the use of the mentioned functions.

A demo of the models is available by running `ddhazard_app`. Particle filter and smoothers are provided with the package but are not covered in this paper. I will end by looking at potential further developments.

1.6.1 Further Developments

I will summarize some potential future developments of the **dynamichazard** in this section. First, we can replace the random walk model with another type of multivariate autoregressive model. This will require additional parameter to be estimated which can be done in the M-step of the EM-algorithm. See the constrained EM-algorithm in the **MARSS** package (Holmes, 2013) for update formulas.

Other models can be implemented in survival analysis, such as recurrent events and competing risk (see Fahrmeir and Wagenpfeil, 1996). Furthermore, the methods can also be used outside survival analysis. For instance, with panel data with real valued outcomes, multinomial outcomes, or ordinal outcomes for each individual in each interval. The underlying time can depend on the context (e.g., it could be calender time or time since enrollment).

The logistic link function in the discrete model can be changed to other link functions without much work as both the C++ and R code is implemented like the `glm` function in R.

The current implementation of parallel computation is based on shared memory. However, we can extend the implementation to a distributed network. Rigatos (2017, chapter 3) covers different ways of performing the computation on a distributed network. Two approaches are either to distribute the work in each step of the filter or to run separate filters and aggregate the filters at the end.

An alternative to the filters in the E-step is to use the linearisation method described in Durbin and Koopman (2012, Section 10.6) mentioned in Section 1.2.4. It would be interesting to implement this approach in the package as well. Fahrmeir and Kaufmann (1991) describe an idea similar to the linearisation method in Durbin and Koopman (2012, Section 10.6), using a Gauss-Newton and Fisher scoring method.

The methods discussed in this paper can be used as the initial input to the importance sampler with use of antithetic variables and control variables, as suggested by Durbin and Koopman (2000). This approach is implemented in the **KFAS** package (Helske, 2017). This can be used for approximate likelihood evaluation to perform maximum likelihood estimation as in the **KFAS** package instead of the EM-algorithm.

All the models covered in this paper can be estimated with a suitable generalized linear mixed model with correlated random terms. Thus, we can perform approximate maximum likelihood estimation with e.g., the pseudo-likelihood method used in the `GLIMMIX` procedure in SAS, or the Laplace approximation used in the `GLIMMIX` procedure and the **lme4** package (Bates et al., 2015) in R. Alternatively, the implemented particle filters in the **dynamichazard** package can be used for approximate likelihood evaluations and parameter estimation.

Chapter 2

Can Machine Learning Models Capture Correlations in Corporate Distresses?

Benjamin Christoffersen, Rastin Matin, and Pia Mølgaard

Abstract

A number of papers document that recent machine learning models outperform traditional corporate distress models in terms of accurately ranking firms by their riskiness. However, it remains unanswered whether advanced machine learning models can capture correlations in distresses sufficiently well to be used for joint modeling, which traditional distress models often struggle with. We implement a regularly top-performing machine learning model and find that prediction accuracy of individual distress probabilities improves while there is almost no difference in the predicted aggregate distress rate relative to traditional distress models. Thus, our findings suggest that complex machine learning models do not eliminate the excess clustering in distresses. Instead, we propose a frailty model, which allows for correlations in distresses, augmented with regression splines. This model demonstrates competitive performance in terms of ranking firms by their riskiness, while providing accurate aggregate risk measures.

Keywords: corporate distress prediction, discrete hazard models, frailty models, gradient boosting
JEL classification: C49, C53, G17, G33

We are grateful to Mads Stenbo Nielsen (discussant), David Lando, Søren Feodor Nielsen, and seminar participants at Copenhagen Business School and Danmarks Nationalbank for helpful comments.

2.1 Introduction

Estimating accurate corporate distress probabilities is of particular interest to central banks in the European Union the coming years. Following the regulation on the collection of credit risk data of the European Central Bank (ECB), members of the euro area are obliged to establish central credit registers and to participate in a joint analytical credit database ("AnaCredit"). The database will contain detailed information on lending by commercial banks to corporate borrowers and, consequently, central banks can closely study the credit risk of a particular bank's corporate loan portfolio. For that purpose, it is essential to model the probability of default of a group of individual borrowers jointly accurate in order to estimate portfolio risk measures.

In this paper, we investigate whether complex statistical models, via their sophisticated dependency structures, can capture correlations in corporate distresses sufficiently well using firm-level data alone. This is motivated by two strands of literature. The first focuses on the application of machine learning models, i.e., complex models with highly nonlinear dependency structures between the covariates and the outcome, to predict corporate bankruptcies (see e.g., Jones et al., 2017, Min and Lee, 2005, Zięba et al., 2016). These papers show applications of one or more complex statistical models which are commonly benchmarked against a logistic regression. Model performance is then evaluated by rank- or binary-based performance metrics that compare the models' ability to classify or predict the distress of a firm. However, the models' ability to accurately estimate the aggregated percentage of firms that will default in the next period remains uninvestigated as well as their ability to provide accurate portfolio risk measures. The second strand of literature, pioneered by Duffie et al. (2009), shows that traditional hazard models (e.g., logistic regression) yield too narrow prediction intervals of the aggregated default rate due to the model assumption that observations are conditionally independent. Duffie et al. (2009) then advocate the need for unobservable temporal effects – or frailty – in the models, which add correlations in defaults after conditioning on covariates, thereby relaxing the conditional independence assumption. The conditional independence assumption is also implicitly made in most complex statistical models. The focus of this paper is to elucidate if this affects such models' ability to accurately estimate the distress rate as well as the risk of a loan portfolio (i.e., not have issues with excess clustering of default).

The complex statistical method we employ is a gradient boosted tree model, which has displayed superior performance in both bankruptcy prediction and other fields.[1] Our hypothesis is that previous models in literature are misspecified due to a linearity and additivity assumption. Violations of these assumptions combined with, e.g., time-varying covariate distributions can yield evidence of excess default clustering. An illustrative example is provided in Appendix 2.D. We do not expect the conditional independence assumption to be satisfied in the gradient boosting model, but our hypothesis is that the effect is sufficiently weak to be practically unimportant.

We find that the gradient boosted tree model is as unable to capture the yearly heterogeneity in distress rates as traditional distress models and, furthermore, it is also unable to provide appropriate estimates for the default risk in a loan portfolio. Comparing results of the gradient boosted tree model to results of a model with frailty, which has closer to nominal coverage of prediction intervals and provides accurate risk measures, we show that loan portfolios of, in particular, large banks can be viewed as too safe in the eyes of the regulator and/or risk manager, if he or she relies on a gradient boosted tree model.

---

[1] See Caruana and Niculescu-Mizil (2006) for a comparison in many other fields and Jones et al. (2017), Zięba et al. (2016) who apply gradient tree boosting to firm distress or bankruptcy prediction with success.

Our sample consists of annual financial accounts published between 2003 and 2016 of all non-financial Danish firms both traded and non-traded. While most of the default literature focuses on public firms where strong market-based predictors are available, private firms are important for the application we have in mind. Private firms have a much larger part of the debt market in Denmark than public firms, as only 14% of the bank debt on the financial statements are held by public firms in our sample at the end of 2016.[2] Moreover, we only have 147 public firms in 2016, which is unrealistic to perform modeling on due to the small number of observed defaults. Thus, we will consider both traded and non-traded firms as this yields a large sample that allows us to include many covariates and add nonlinear effects. The work in the main body of the paper is based solely on micro level data, but in a robustness test we show that models including macro level data perform better in some periods. However, estimating a model that generalizes well may be hard with the limited number of cross-sections. Lastly, the unobserved temporal effect is still economically and statistically significant after the inclusion of the macro variable.

We start the analysis by benchmarking the gradient boosted tree model against a multiperiod logit model as in Shumway (2001) and a generalized additive model, which allows for a nonlinear relationship between the covariates and the probability of entering into a distress on the logit scale. Like others before us, we observe improvements in out-of-sample ranking of firms by their distress probability as we use more complex models, going from an average out-of-sample area under the receiver operating characteristic curve (AUC) of 0.798 to 0.822. Thus, we find that the more complex model is 2.4 percentage points more likely to have a higher distress probability for a random distressed firm than for a random non-distressed firm in each year on average. However, the gains we find of complex modeling are more than 4 times smaller than what recent papers find.[3] Thus, one may prefer the simpler models if interpretability is desired with only a minor loss of accuracy. Our finding suggests that earlier papers have used poor baseline models when evaluating the gains of applying complex machine learning models. Further, the difference between the firm-level performance of our generalized additive model and gradient boosted tree model is small. This result suggests that higher order interactions may not be needed for corporate default models, as our generalized additive model only allows for two-way interactions.

Next, we address the models' ability to predict the percentage of firms that will enter into a distress in the following period. We find that all models fail to capture the temporal fluctuations in distress rates and provide too narrow prediction intervals. In particular, only very few of the 90% prediction intervals contain the realized percentage of firms entering into distress in the 10 years that we can backtest. We formally test the models' ability to provide accurate prediction intervals by backtesting estimated value-at-risk like figures of the distress rates for different portfolios that mimic bank exposures. All three models fail the test at a 1% significance level with a null hypothesis that the quantiles have the correct coverage. Thus, none of the models have wide enough prediction intervals or provide accurate risk measures.

---

[2] The figures is computed by taking the bank debt provided by Bisnode and subtracting the bond debt which is included in these figures.

[3] See Zięba et al. (2016) and Jones et al. (2017).

Too narrow prediction bounds have several implications. First, they may result in a downward bias in risk measures for a portfolio with exposures to different firms. Secondly, they suggest that the assumption of conditional independence given the covariates is not satisfied. Violation of the conditional independence assumption suggests that there may exist an unobservable macro effect that needs to be accounted for to capture the excess distress events. That is, the gradient boosted tree model is not sufficiently able to capture correlations in distresses from firm-level data alone despite fitting better to each individual firm. The fact that the conditional independence assumption is violated may not be surprising, but the violation is sufficiently large that it cannot be disregarded.

To relax the conditional independence assumption we estimate a generalized linear mixed model (a frailty model) with a random intercept which allows for correlations in distresses beyond the correlation introduced by the covariates. We contribute to the current literature on frailty models by adding nonlinear effects between the covariates and the outcome variable. Inclusion of nonlinear effects yields a better firm-level model and we thus get a frailty model which provides out-of-sample rankings that are almost as good as the gradient boosted tree model. Moreover, we show that the random intercept in the frailty model is both statistically and economically significant.

2.2 Related Literature

This paper combines two strands of literature in the field of predicting corporate defaults. The first focuses on frailty (and/or contagion) or time-varying effects (e.g., see Azizpour et al., 2018, Chen and Wu, 2014, Duffie et al., 2009, Koopman et al., 2011, Kwon and Lee, 2018, Lando et al., 2013, Qi et al., 2014). These papers generally show that models with a simple relationship between observable covariates and distress fail to capture the yearly fluctuations in default rates, i.e., a violation of the conditional independence assumption. Various forms of unobservable effects are then introduced which account for the temporal fluctuations. Our contribution to this line of work is a frailty model, where nonlinear dependencies are introduced between some covariates and the outcome variable on the log odds scale. Furthermore, we compare the frailty model to a statistical model that allows for complex dependencies between covariates and the outcome variable and find that the frailty model shows almost as good ranking and better coverage of the prediction intervals. Moreover, we provide evidence that the excess clustering of defaults is not because of a too simple dependency structure with a large sample of private and public firms.

The second strand of literature that we relate to applies complex statistical models to improve probability of default estimates (e.g., see Jones et al., 2017, Kim and Kang, 2010, Min and Lee, 2005, Zięba et al., 2016). These papers often use considerably more covariates in their models and use methods which allow for more complex relationships compared to typical frailty models. The main focus of these papers is on ranking or binary classification of firms and not on whether the models capture the temporal fluctuations in the default rate. The complex models are typically benchmarked against a logistic regression (among other models) with automated model selection and little focus on model diagnostics. In our paper we use a logistic model as a benchmark as well, but we carefully set up the model using both statistical and economic sense. We add to this literature by evaluating the ability of the complex model to capture the yearly fluctuation in default rates. We show that the improvements in the forecasts for each firm do not outweigh the strict conditional independence assumption when one is interested in portfolio risk.

2.3 Statistical Models for Predicting Corporate Distress

In this section we go through the four discrete hazard models used in this paper to predict corporate distress. The discrete hazard models we use are estimated using a panel data set where each observation contains a set of covariates (financial ratios, age, sector etc.) and an indicator of whether the firm has a distress event or not in the given year. We will cover the distress event definition and discrete hazard models further in Section 2.4.1.

First, we briefly describe the well known multiperiod logit model. The notation introduced in this section will serve as the basis for the more general models. Secondly, we describe the generalized additive model which allows for a nonlinear dependence between the covariates and the probability of distress on the logit scale. Thirdly, we describe the gradient boosted tree method we use. Finally, we introduce the generalized linear mixed model which relaxes the conditional independence assumption.

2.3.1 Generalized Linear Models

We will use so-called multiperiod logit models, where we employ a logistic regression in the discrete hazard model described in Section 2.4.1. Let $R_t \subseteq \{1, \cdots, n\}$ denote the active firms at time t, y_{it} denote whether firm i has an event in year t, d denote the number of years, and $\boldsymbol{x}_{it}$ denote the covariates for firm i in year t. Then the maximum likelihood estimates of the coefficients, $\boldsymbol{\beta}$, are

$$\arg\max_{\boldsymbol{\beta}} \sum_{t=1}^{d} \sum_{i \in R_t} y_{it} \boldsymbol{\beta}^\top \boldsymbol{x}_{it} - \log\left(1 + \exp\left(\boldsymbol{\beta}^\top \boldsymbol{x}_{it}\right)\right) \tag{2.1}$$

where $\boldsymbol{x}_{it}$ includes a constant 1 for the intercept, industry covariates, and potentially macro covariates. Furthermore, we will refer to R_t as the *risk set* and let $\mathbf{X}_t$ denote the matrix with rows equal to the covariate vectors $\boldsymbol{x}_{it}$ with $i \in R_t$. Multiperiod logit models are a common choice for distress models since the work of Shumway (see Shumway, 2001) and has been used in, e.g., Campbell et al. (2008), Chava and Jarrow (2004). We will refer to multiperiod logit models as generalized linear models (GLMs) since estimation can be done with regular estimation methods for GLMs.

2.3.2 Generalized Additive Models

The GLM in Section 2.3.1 may pose too strict assumptions on the relationship between the covariates and whether a firm enters into distress. In particular, the assumption that the covariates are linearly related to the logit of the probability of distress may be too strict for some of the covariates. Generalized additive models (GAMs) relax this assumption by assuming that some of the covariates have a continuous and nonlinear partial effect on distress probability on the logit scale.

We employ a GAM where nonlinear effects are accounted for through natural cubic splines with a penalty on the second order derivative. The maximization problem with q nonlinear effects and with given penalty parameters $\boldsymbol{\lambda} = (\lambda_1, \dots, \lambda_q)^\top$ is

$$\boldsymbol{\beta}(\boldsymbol{\lambda}) = \arg\max_{\boldsymbol{\beta}} \sum_{t=1}^{d} \sum_{i \in R_t} y_{it}\eta_{it} - \log\left(1 + \exp\left(\eta_{it}\right)\right) - \boldsymbol{\beta}^{(s)\top}\mathbf{S}(\boldsymbol{\lambda})\boldsymbol{\beta}^{(s)} \tag{2.2}$$

where

$$\eta_{it} = \boldsymbol{\beta}^{(f)\top}\boldsymbol{x}_{it}^{(f)} + \sum_{j=1}^{q} \boldsymbol{\gamma}_j^\top \boldsymbol{f}_j(x_{itj}^{(s)}), \quad \boldsymbol{\beta}^{(s)} = \begin{pmatrix} \gamma_1 \\ \vdots \\ \gamma_q \end{pmatrix}, \quad \boldsymbol{\beta} = \begin{pmatrix} \boldsymbol{\beta}^{(f)} \\ \boldsymbol{\beta}^{(s)} \end{pmatrix} \tag{2.3}$$

The functions $\boldsymbol{f}_j$ return a basis vector for a natural cubic spline, $x_{itj}^{(s)}$ is covariate j of firm i with a nonlinear effect at time t, $\boldsymbol{x}_{it}^{(f)}$ are the covariates with a linear effect for firm i at time t, and $\mathbf{S}(\boldsymbol{\lambda})$ is a penalty coefficient matrix which yields a second order penalty on each spline $j = 1, \dots, q$. The knots for the natural cubic spline basis are chosen as empirical quantiles. Equation (2.2) can be solved with penalized iteratively re-weighted least squares if $\boldsymbol{\lambda}$ is known.

The penalty coefficient matrix, $\mathbf{S}(\boldsymbol{\lambda})$, depends linearly on the unknown penalty parameters $\boldsymbol{\lambda}$ that have to be estimated. This is done by minimizing an un-biased risk estimator (UBRE). The effective degrees of freedom is the trace of

$$\mathbf{F}_{\boldsymbol{\lambda}} = \left(\mathbf{X}^\top\mathbf{W}\mathbf{X} + \mathbf{S}(\boldsymbol{\lambda})\right)^{-1}\mathbf{X}^\top\mathbf{W}\mathbf{X} \tag{2.4}$$

where $\mathrm{tr}\left(\cdot\right)$ denotes the trace of a matrix and $\boldsymbol{y}$, $\mathbf{X}$, $\boldsymbol{z}$, and $\mathbf{W}$ denote the stacked matrices and vector from each year (e.g., $\boldsymbol{y} = (\boldsymbol{y}_1^\top, \dots, \boldsymbol{y}_d^\top)^\top$). The columns of $\mathbf{X}$ include the evaluated basis functions, $\boldsymbol{f}_j$, for the nonlinear effects. $\mathbf{W}$ and $\boldsymbol{z}$ are the diagonal matrix with working weights and vector of pseudo-responses from iterative re-weighted least squares, respectively. They implicitly depend on $\boldsymbol{\beta}(\boldsymbol{\lambda})$, $\boldsymbol{y}$, and $\mathbf{X}$.[4] The maximization is done with the so-called performance-oriented iteration. See Wood (2017), Wood et al. (2015) for further technical details.

The final model also includes tensor product splines to allow for smooths in two dimensions. These are formed by taking an outer product of two spline basis functions $\boldsymbol{f}_j$ and is more general than the model in Equation (2.3). The extension to two-dimensional smooths is straightforward, but not covered in Equation (2.3) to keep the notation simple. GAMs have received limited attention in the corporate default literature (one example is Berg, 2007). This is despite that there is no prior reason to expect that the associations with covariates should be linear and additive on the logit scale. While we may expect a monotone effect, the linearity and additivity are not obvious.

---

[4] Let g denote the link function which maps from the probability of an event to the linear predictors, η_{it}, in Equation (2.3), let $\widehat{p}_{it} = g^{-1}(\eta_{it})$ be the expected probability of an event at the current iteration during estimation or at convergence, and let $V(p) = p(1-p)$ denote the map from the probability of an event to the variance. Then $z_{it} = \eta_{it} + g'(\widehat{p}_{it})(y_{it} - \widehat{p}_{it})$ and $w_{iit} = 1/g'(\widehat{p}_{it})^2 V(\widehat{p}_{it})$.

The advantage of the GAM is that the researcher has control over the complexity of the model. For example, he or she can decide which covariates should be modeled with a nonlinear effect and which do not. Moreover, it is easy to validate whether the final model makes sense through standard diagnostic plots, to obtain partial effects of the covariates, to compute confidence intervals, etc. We illustrate this in Appendix 2.B.1. However, the researcher has to consider the effect and interactions of the covariates which may be hard, especially with higher order nonlinear interactions.

Our selection of nonlinear effects and interactions is data driven. Particularly, we use standard residual plots to see deviations from linearity and add nonlinear effects accordingly. While one may be concerned about overfitting with a data driven procedure then the splines we show in Appendix 2.B.1 have very low uncertainty in areas with a large number of observations and observed events. Although, these are computed under the assumption of conditional independence, our results would not be altered by a substantial increase in the uncertainty in these areas.

2.3.3 Gradient Tree Boosting

Gradient tree boosting (GB) is a greedy function approximation method that can approximate very complex models. The method is greedy in the sense that we iteratively make small local improvements without updating the previous parts of the model. GB has gained much attention possibly due to its flexibility and easy usability, as the researcher only has a few and simple model choices relative to the GAM described in Section 2.3.2. Furthermore, GB has shown superior performance in many fields, see e.g., Caruana and Niculescu-Mizil (2006), where an empirical study is presented on different data sets where GB performs best on average on many metrics. However, the advantages of GB come at a cost of limiting the researcher's ability to set the complexity of the effect of each covariate. Furthermore, it is not clear how to perform inference such as testing significance of partial effects, and evaluating if the final model is “sensible” for various combinations of covariates may be difficult if one allows for higher-order interactions (i.e., deep trees). Lastly, figuring out why a given observation gets the predicted probability is not as easily done as with the GLM and GAM. This is a drawback for a financial institution that is required to provide an explanation of why a certain probability of distress is predicted.

We will cover gradient tree boosting in the context of classification with the logit link function.[5] We use Newton boosting, but in the following we will refer to it as gradient boosting as commonly done in literature. A key component in the GB model is regression trees. The regression trees we use solve a weighted L2 minimization problem by repeatedly performing binary splits on one of the covariates. The splits are found greedily by taking the covariate and splitting point (of the points that are considered) which yields the best improvements at each iteration. Figure 2.1 shows the first regression tree of the GB model estimated on the full sample.

---

[5] We refer to Bühlmann and Hothorn (2007), Friedman (2001) and Chen and Guestrin (2016) for more details regarding the method and the software implementation of GB that we use, respectively.

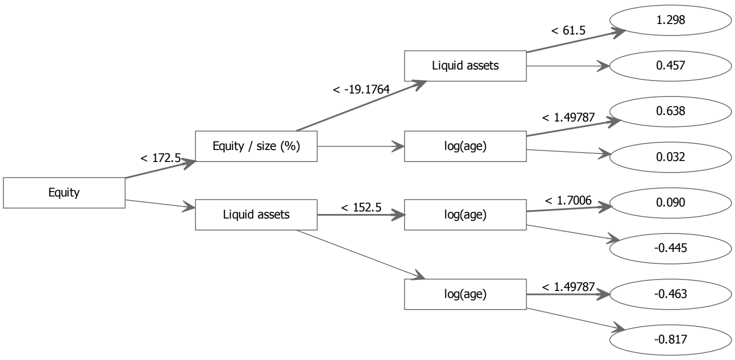


Figure 2.1: Example of a regression tree. The figure shows the first tree in a GB model estimated on the full sample. Squares are parent nodes while ovals are leaves. The figures in the leaf nodes are the log odds value which is the term added to the linear predictor (multiplied by a shrinkage parameter) in the final GB model. Observations in the top leaf are the most risky. These have negative equity and a low value of liquid assets. Observations in the bottom leaf are the least risky firms with high equity, high liquid assets, and are older firms.

We now turn to gradient boosting. Denote the estimated mean probability of a distress by

$$\bar{p} = \frac{1}{n} \sum_{t=1}^{d} \sum_{i \in R_t} y_{it}$$

where $n_t = |R_t|$ is the number of active firms at time t and $n = \sum_{t=1}^{d} n_t$ is the total number of observations. Initialize the linear predictors as $\eta_{it}^{(0)} = f^{(0)}(\boldsymbol{x}) = \text{logit}(\bar{p})$, where $\text{logit}(p) = \log(p/(1-p))$ is the logit function. Let $\mathbf{X}$, $\boldsymbol{y}$, and $\boldsymbol{\eta}^{(i)}$ denote the stacked matrix and vectors such that, e.g., $\boldsymbol{\eta}^{(i)} = (\boldsymbol{\eta}_1^{(i)\top}, \dots, \boldsymbol{\eta}_d^{(i)\top})^\top$. Define the loss function, L, as

$$\begin{aligned} L(\boldsymbol{\eta}) &= \sum_{t=1}^{d} \sum_{i \in R_t} l(\eta_{it}; y_{it}) \\ l(\eta; y) &= -y\eta + \log\left(1 + \exp(\eta)\right) \end{aligned}$$

Then for $i = 1, \dots, k$

1. compute the first and second order derivatives using the linear predictors from the previous iteration and denote these by

$$\begin{aligned} g_{it} &= -y_{it} + \left(1 + \exp\left(-\eta_{it}^{(i-1)}\right)\right)^{-1} \\ h_{it} &= \exp\left(-\eta_{it}^{(i-1)}\right)\left(1 + \exp\left(-\eta_{it}^{(i-1)}\right)\right)^{-2} \end{aligned}$$

2. fit a regression tree denoted by $a^{(i)}(\boldsymbol{x})$ which is an approximation to

$$\underset{a \in \mathcal{C}}{\arg\min} \sum_{t=1}^{d} \sum_{i \in R_t} h_{it} \left(-\frac{g_{it}}{h_{it}} - a(\boldsymbol{x}_{it})\right)^2$$

where $\mathcal{C}$ is the set of trees we consider (e.g., trees with a given maximum depth).

3. update the model such that $f^{(i)}(\boldsymbol{x}) = f^{(i-1)}(\boldsymbol{x}) + \rho a^{(i)}(\boldsymbol{x})$, where $\rho \in (0, 1]$ is a predetermined shrinkage parameter.
4. update the linear predictors by computing $\eta_{it}^{(i)} = f^{(i)}(\boldsymbol{x}_{it})$.

The final GB model is the function $f^{(k)}$. The greedy part of the procedure can be seen at step 2 and 3 where update model locally without changing the previous parts of the model. There are three main parameters in the above algorithm: The shrinkage parameter ρ, the maximum depth of the trees in step 2, and the number of trees k. We select these with 5-fold cross-validation where we sample the firms (not the financial statements) and evaluate the AUC which is introduced in Section 2.5. In general, it is preferable to decrease the shrinkage parameter, ρ, while increasing the number of trees, k, to get a better approximation of the true dynamics. However, one has a finite budget in terms of computational power, which limits k and thus forces one to select ρ to get the optimal number of trees around k. We fix k to around 250 and at most 300. We find ρ with cross-validation on the full sample from 2003-2016 which we will describe in Section 2.4.1. We find only very small improvements of decreasing the learning rate and using more trees. This only leaves us with a choice for the maximum depth of trees.

Usually so-called 'weak learners' (biased methods) are used in step 2 above. In our case, this amounts to shallow trees (trees with a low maximum depth). The weak learners are then combined through gradient boosting yielding one "good" model with a substantially lower bias than any of the individual learners while not affecting the variance much. See Bühlmann and Hothorn (2007) for some simpler examples with theoretical results. For the aforementioned reasons, we have tried maximum depths of 2-6 in preliminary testing. We used 5-fold cross-validation as described above. We find little difference in model performance when going from tree depths of 3 to 6 and, thus, we choose a maximum depth of 3.

Given the fixed learning rate and maximum depth of 3, we estimate the optimal number of trees each year when we run our out-of-sample tests. The estimations are done again with 5-fold cross-validation and by sampling firms and not financial statements. We note that the estimation of the optimal number of trees is done on the estimation sample and not the test set. As for our main sample, we do not expect that decreasing the learning rate and increasing the number of tress will have an impact on our results.

2.3.4 Generalized Linear Mixed Models

We can extend the GLM from Section 2.3.1 to relax the conditional independence assumption by generalizing to a generalized linear mixed model (GLMM). This can be done by changing the conditional mean in the GLM from

$$\mathrm{E}\left(Y_{it} \mid \boldsymbol{x}_{it}\right) = g^{-1}(\boldsymbol{\beta}^\top \boldsymbol{x}_{it}), \quad g^{-1}(\eta) = \operatorname{logit}^{-1}(\eta)$$

to

$$\mathrm{E}\left(Y_{it} \mid \boldsymbol{x}_{it}, \boldsymbol{z}_{it}, \epsilon_t\right) = g^{-1}\left(\boldsymbol{\beta}^\top \boldsymbol{x}_{it} + \epsilon_t\right) \tag{2.5}$$

where $\epsilon_t \sim N\left(0, \sigma^2\right)$ is the random effect at time t and $\epsilon_t \perp \epsilon_s$ for $t \neq s$. Thus, the optimization problem becomes

$$\underset{\boldsymbol{\beta}, \sigma^2}{\arg\max} \sum_{t=1}^{d} \int_{\mathbb{R}} \left(\sum_{i \in R_t} y_{it}\eta_{it} - \log\left(1 + \exp\left(\eta_{it}\right)\right) \right) \varphi\left(\epsilon_t; \sigma^2\right) \partial\epsilon_t \tag{2.6}$$

$$\eta_{it} = \boldsymbol{\beta}^\top \boldsymbol{x}_{it} + \epsilon_t \tag{2.7}$$

where $\varphi(x, \sigma^2)$ is the density function of a normal distribution with zero mean and variance σ^2. The log-likelihood in Equation (2.6) has no closed form solution in general, but can be approximated with a Laplace approximation. Furthermore, the computational cost of the approximation can be greatly reduced if one exploits the sparsity of the matrices which are decomposed during the estimation. See Bates et al. (2015) and the citations therein for further details about the estimation method. The linear predictor in Equation (2.7) is easily modified to include splines by changing the $\boldsymbol{\beta}^\top \boldsymbol{x}_{it}$-term such that

$$\eta_{it} = \boldsymbol{\beta}^{(f)\top} \boldsymbol{x}_{it}^{(f)} + \sum_{j=1}^{q} \boldsymbol{\gamma}_j^\top \boldsymbol{f}_j(x_{itj}^{(s)}) + \epsilon_t, \quad \boldsymbol{\beta}^{(s)} = \begin{pmatrix} \gamma_1 \\ \vdots \\ \gamma_q \end{pmatrix}, \quad \boldsymbol{\beta} = \begin{pmatrix} \boldsymbol{\beta}^{(f)} \\ \boldsymbol{\beta}^{(s)} \end{pmatrix}$$

which is similar to Equation (2.3). We denote the random effect, ϵ_t, as frailty though it is not a frailty in the original sense as in Vaupel et al. (1979). The random effect variable in Vaupel et al. (1979) and Duffie et al. (2009) is a multiplicative factor on the hazard. Our random effect is multiplicative on the odds rather than the hazard since we can factorize Equation (2.5) when g is the logit function as

$$\frac{\mathrm{E}\left(Y_{it} \mid \boldsymbol{x}_{it}, \boldsymbol{z}_{it}, \epsilon_t\right)}{1 - \mathrm{E}\left(Y_{it} \mid \boldsymbol{x}_{it}, \boldsymbol{z}_{it}, \epsilon_t\right)} = \exp\left(\boldsymbol{\beta}^\top \boldsymbol{x}_{it}\right) \exp\left(\epsilon_t\right)$$

Thus, firms are more "frail" if ϵ_t is large in a given year, yielding an $\exp(\epsilon_t)$-factor higher odds of distress. The case $\epsilon_t = 0$ can be seen as a "standard" year. Random effect models have received a lot of attention in the literature, where the focus is on the structure of the random effects. E.g., Duffie et al. (2009)[6] and Koopman et al. (2011) let the probability of distress depend on an unobservable order-one autoregressive process. However, contrary to Duffie et al. (2009), Koopman et al. (2011) assume that groups of firms depend differently on the unobservable process. We are limited in terms of how many random effects we can estimate as we only have 14 years of data. Thus, we will only estimate a single random intercept, where we assume that the ϵ_t-terms are iid as in Equation (2.6). An autocorrelation plot of predicted ϵ_t-terms does not show signs of autocorrelation.

---

[6] Duffie et al. (2009) use an Ornstein–Uhlenbeck process for the random intercept term but use a discrete approximation in which case one gets an order-one autoregressive process.

2.4 Data and Event Definition

Our main data set consists of all non-consolidated financial statements filed by Danish private and public limited companies in the period 2003 to 2016.[7] The financial statements are supplemented with firm characteristics such as age, sector, and legal status from the Danish Central Business Register (CVR). As we are conducting a prediction exercise, we utilize financial statements as of their publication date and not the accounting period end date. In our sample, financial reports are typically made public 5 months after the accounting period end date,[8] where we only use the most recent published accounting data for each firm from year $t-1$ in year t in our models.

We apply standard filters to focus the analysis on the core of the Danish corporate sector. First, we exclude financial firms and holding companies as is typically done in the literature. Further, we exclude a small number of financial statements which are filed in Denmark in other currencies than DKK, EUR, GBR, USD, or SEK.[9] We do not impose any restrictions on firm size as we want to capture the whole economy in the analysis.

It could be argued that the analysis should only focus on "large" firms, as they hold the majority of the total assets and debt. Instead of estimating models on just large firms, we allow for interactions between firm size and other variables in the GAM and GB models, thereby creating different models for firms of different sizes. Among the interactions tested we find that the interaction between scaled net profit and the log of the size variable we introduce later as well as between scaled liquid assets and log size are significant in the GAM. Furthermore, including small firms increases our sample size, which is important in order to estimate the nonlinear effects in the GAM and GB model. The performance with respect to large firms is improved when we estimate a separate GLM for large firms, but remains inferior to the performance of GAM and GB model. For consistency, all results will be of the models estimated on the full sample.

The filtered data set includes 198 929 individual firms and 1.3 million firm years in the 2003 to 2016 period. Of the 198 929 firms, 43 674 enter into a distress period at least once. The seemingly high rate is due to a larger distress rate for small firms. An interesting aspect of our sample is that it includes non-traded firms, which are less studied in the literature. Models for private firms are particularly relevant because of the large fraction of bank debt held by private firms.

2.4.1 Event Definition and Censoring

We obtain information on the full history of each firm's status from the Danish Central Business Registry (CVR). The CVR categorizes firm status into 21 categories. We combine categories into three groups: “normal”, “in distress” and “other”.[10] The “in distress” category includes firms in bankruptcy, firms that went bankrupt, firms under compulsory dissolution, or firms that have ceased to exist due to compulsory dissolution.

---

[7] Financial statements are delivered to us by Bisnode and Experian.

[8] The exact publication date is used for statements filed from 2012 and onward. Unfortunately, we do not have access to the publication date of statements filed before 2012. For these statements we set the publication date to 6 months after the accounting period end date. We have two reasons for doing this. First, Danish law requires that the majority of firms in our sample must publish their financial statements within 5 months of the accounting period end date. We use 6 months instead of 5 to be conservative. Secondly, we find that 96% of financial statements are published within 6 months of the accounting period end date in the sub-sample where we have the publication date.

[9] Accounting variables reported in other currencies than DKK are converted to DKK as follows. All stock variables are converted using the end of accounting period exchange rate. All flow variables are converted using the daily average exchange rate over the accounting period.

Our definition of “in distress” implies that firms that are “in distress” can become active again. Thus, we model recurrent events. We choose this framework as creditors are likely to suffer losses when a firm enters into a distress period, even if the firm becomes active again, due to delayed payments or a write-down of the debt. In our sample 3.4% of the firms have experienced a prior distress (some before 2003) and have recovered and, furthermore, 1 352 of these firms enter into more than one distress period during our sample period.

Distress dates are highly seasonal and reflect a potentially delayed processing time of the authorities.[11] Thus, we limit the models to be on a yearly basis. Each year includes all firms that:

1. had a “normal” status at the end of the previous year.
2. published a financial statement within the previous year.
3. (a) enter into “in distress” the following year or
   (b) do not publish a new financial statement the following year and enter into the “in distress” status within two years of the publication date of the latest financial statement or
   (c) are still “normal” at the end of the year (i.e., are not censored).

Firms that fulfil all of the above conditions are denoted *active* at the beginning of the given year. Among these firms, we say that a firm has an *event* if it satisfies condition 3a or 3b, or that the firm is a *control* if it satisfies condition 3c. Condition 3b is similar to the event definition in Shumway (2001), who defines a firm as going bankrupt if the firm delists the following year and “files for any type of bankruptcy within 5 years of delisting”. The difference to our data set is that firms do not delist, but instead do not publish a new financial statement. We also include a few firms that satisfy 3a or 3b as events if they enter into the “other” status between the “normal” and the “in distress” status.

In our event definition we have chosen a window of 2 years between the publication date of the last financial statement and the declaration date of “in distress”. Most distresses in our sample are declared approximately 1.5 years after the publication date of the last financial statement but some occur later. We find, across years, that 95% to 99% of all “in distress” statuses are declared within the 2 year window we have chosen.

2.4.2 Covariates

It is common to scale most of the financial statement variables by total assets to get all financial statement variables on a common scale and such that they are reasonable to include on the log odds scale. However, a non-trivial fraction of the firms in our sample have negative equity at some point. Thus, using total assets as the denominator will yield extreme covariates which may not fit well in a GLM. As Campbell et al. (2008) we define a more suitable metric to capture the firm size. We define firm size as

---

[10] The “other” group includes firms that are under liquidation, liquidated, merged and split.

[11] Every year, there is a large “peak” in reported distress events in a single month in the fall and this peak does not fall on the same month every year. This arbitrary peak in reported distress events makes it questionable whether there is any meaning in the exact reporting month.

$$\text{size}_{it} = \max\{\text{debt}_{it}, \text{total assets}_{it}\} \tag{2.8}$$

where debt_{it} and total assets_{it} refer to the debt and total assets of firm i on the balance sheet from the financial statement published between year $t-1$ and t, respectively. Thus, size_{it} equals the total debt of the firm when equity is negative and otherwise total assets. We use size_{it} in the denominator of all the ratios where we would otherwise use total assets.

Besides the financial statement variables we include some variables that we have constructed ourself. Most interestingly, we include an industry-specific covariate as in Chava et al. (2011) by computing the average net profit divided by the size variable each year for each leading four digit standard industrial classification (SIC) group. Unlike Chava et al. (2011), we do not have the stock return so we use the net profit divided by the size variable. Moreover, Chava et al. (2011) include a dummy for whether median stock return in the industry is below -20%. We do not believe that the variable has a discrete effect upon exceeding a pre-specified threshold and, therefore, we include the average value and estimate a slope. We winsorize[12] all covariates at the 5%- and 95%-quantile as in Campbell et al. (2008) since preliminary results showed influential observations and poorer fits in the GLM when more extreme quantiles were used.

We end up with 44 numerical and 6 categorical covariates. We use the Thresholded Lasso estimator described in Appendix 2.A to select the covariates we will use in the GLM, GAM, and GLMM. This is similar to Tian et al. (2015) except that we have an additional step in our algorithm to deal with bias issues with the Lasso estimator. We exclude 3 of the covariates, but include them all in the GB model as the GB model tends to be robust against redundant covariates (e.g., this model does not have issues with multicollinearity as the other models). Another advantage of the GB model is that the regression trees used in the model are invariant to monotone transformations of the covariates. Consequently, we include both the non-winsorized ratios and the original (non-ratio) figures from the financial statements in the GB model. Descriptive statistics of all of the covariates can be seen in Appendix 2.A.

2.5 Performance of the GLM, the GAM, and the GB Model

In this section we perform out-of-sample tests of the GLM, GAM, and GB model presented in Section 2.3.1, 2.3.2, and 2.3.3, respectively. We will use an expanding window of data to estimate the models and forecast the probability of the firms entering into distress two years after the estimation window closes. As an example, we use models estimated on 2003 to 2007 to predict default probabilities in 2009. The two-year gap mimics the true forecasting situation as the definition of the distress event requires a lag of two years.

---

[12] Cap values at a given high level quantile and floor at a given low level quantile. We winsorize ratios and not the numerator and denominator separately for ratio covariates.

We measure performance on several different metrics. First, we consider the accuracy of the individual probability-of-distress estimates by comparing the AUC and the log score of the individual models. Next, we consider the performance of the models at an aggregated level by examining the models' ability to predict next year's aggregated percentage of firms in distress as well as the fraction of debt in distress. Finally, we look at the models' ability to estimate portfolio risk.

In-sample results on the 2003 to 2016 data set are presented in Appendix 2.B. The appendix also includes some details of the final model specifications, illustrations of the estimated models, and comparisons between the models. The in-sample results are left as an appendix to allow the paper to focus on the forecasting ability of the models.

2.5.1 Evaluating Individual Distress Probabilities

We start by evaluating the models by their respective AUC. The AUC is a commonly used metric to evaluate out-of-sample performance. It measures the probability that a model places a higher event probability on a random firm that experiences an event in a given year than a random firm that does not experience an event in a given year. Hence, 0.5 is random guessing and 1 is a perfect result.

Figure 2.2(a) shows the out-of-sample AUCs. In all years we find that the GB model gives the highest AUC and therefore is best at ranking firms by their distress risk, followed by the GAM and the GLM. This observation is consistent with the findings in Zięba et al. (2016) and Jones et al. (2017) in the sense that they also find that GB models are superior in terms of AUC. However, the differences we measure in AUCs are much smaller than reported in the aforementioned papers. We find that the average AUC across years are 0.798, 0.811, and 0.822 for the GLM, GAM, and GB model, respectively. Hence, there is an improvement in AUC between the GLM and the GB model of only 0.024. In comparison, Zięba et al. (2016) and Jones et al. (2017) find improvements in the AUC between a benchmark logistic regression and boosted tree model above 0.1. We reckon that the greater improvement in AUC is, to a large extent, due to the GLM used in Zięba et al. (2016) and Jones et al. (2017).[13]

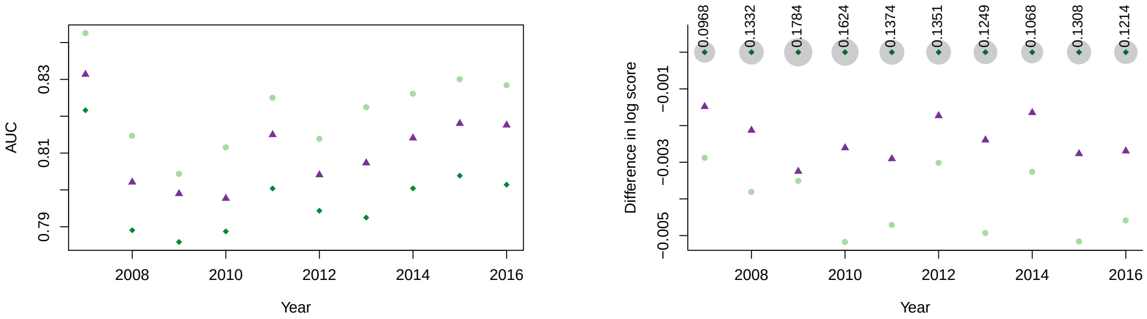
As mentioned above, the AUC is only a ranking measure. A model may rank the firms well, but perform poorly in terms of the level of the predicted probabilities. As we are also interested in well-calibrated probabilities, we look at the log score which is computed by

$$\mathcal{L}_{tj} = -\frac{1}{n_t} \sum_{i \in R_t} \left(y_{it} \log \left(\widehat{p}_{itj} \right) + (1 - y_{it}) \log \left(1 - \widehat{p}_{itj} \right) \right) \tag{2.9}$$

where $\mathcal{L}_{jt}$ is the log score of model j in year t, y_{it} is a dummy equal to 1 if firm i has an event in year t, $\widehat{p}_{itj}$ is the predicted probability of distress of firm i in year t by model j, R_t is the sample of active firms in year t, and n_t is the number of firms in R_t. A perfect score is zero.

---

[13] The data set used in Zięba et al. (2016) is publicly available. We can confirm that the results for the GLM can be greatly improved with limited effort. Both cited papers use raw accounting figures like "Annual Growth in Capital Expenditure" and total assets without any transformations. While these may work well in tree algorithms which are invariant to monotonic transformations then it seems very unseasonable to assume a linear association on the log odds scale as they do in their logistic regression model.



(a) Area under the ROC curve

(b) Log score

Figure 2.2: More complex models have higher AUC and better log scores. The figure shows performance measures of the three models (GLM ♦; GAM ▲; GB ●). Panel (a) shows out-of-sample area under the receiver operating characteristics curve (AUC) for the different models. Panel (b) illustrates the out-of-sample log scores of the three models. The figures above the center of the grey circles are the log scores for the GLM and the areas of the circles are proportional to the figures. The points show the log score of the model minus the log score of the GLM. That is, $\mathcal{L}_{tj} - \mathcal{L}_{t\text{GLM}}$ where $\mathcal{L}_{tj}$ is defined in Equation (2.9) and $j \in \{\text{GLM}, \text{GAM}, \text{GB model}\}$. The models are estimated on an expanding window of data with a 2-year gap to the forecasted data set. E.g., the models which are used to forecast the 2011 distresses are estimated on 2003-2009 data.

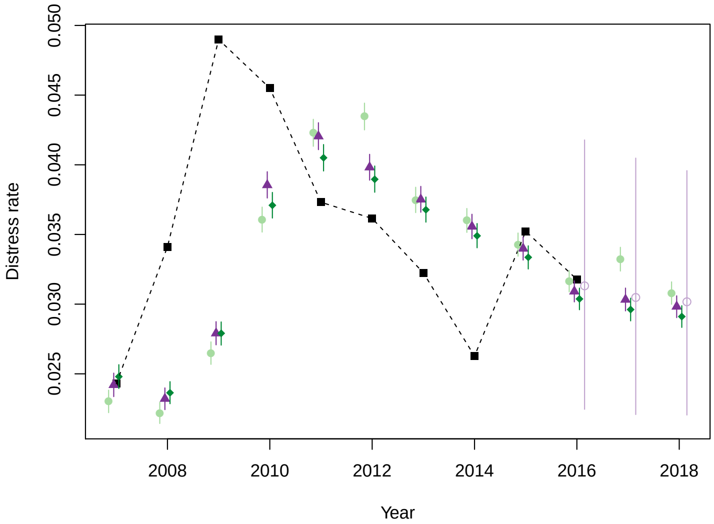
The out-of-sample log scores are illustrated in Figure 2.2(b), where we observe that the GB model outperforms the other models for all years. However, as with the AUC, we find that the improvements in log score with more complex models are relatively small.

To summarize, we find evidence that the GB model is the best model at estimating individual default probabilities. However, the improvements are not large compared to the GAM. Thus, one may prefer the GAM model if interpretability is important.

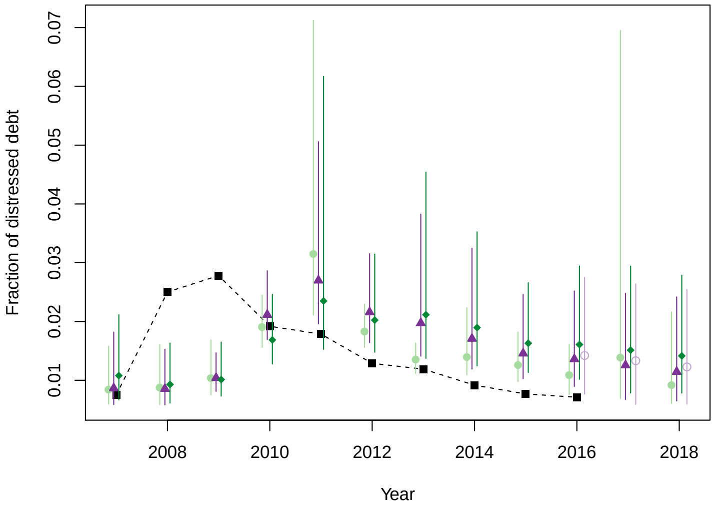
2.5.2 Evaluating Aggregated Distress Probabilities

In this section, we look at the models' ability to predict the distress risk of the aggregated sample. Figure 2.3(a) shows the realized percentage of firms entering into distress as well as the out-of-sample predicted percentage of firms that will enter into distress each year for each of the models. All four models are included in the figure for later comparison, but for now we will only discuss results of the GLM, GAM, and GB model. It is clear that none of the models capture the distress level. Furthermore, none of the models' 90% prediction intervals have close to 90% coverage, which indicates that the assumed conditional independence assumption is violated, i.e., there is some correlation in distress events which is not accounted for in any of the models. That is, the complex GB model is just as bad at capturing the aggregated distress level as the more simplistic GLM. We run a formal test of the models' ability to estimate risk measures in Section 2.5.3.

The aggregated distress rate of the GB model in 2012 and 2017 is higher and in 2012 further away from the realized value than the distress rate of the other models. This raises the question whether the GB model suffers from overfitting, which is not the case as we use cross-validation to select the number of trees. Furthermore, the out-of-sample aggregate distress rates of the



(a) Realized and predicted distress rate



(b) Realized and predicted fraction of debt in distress

Figure 2.3: Models without frailty are unable to predict aggregated distress levels. The figures compare realized percentage of firms in distress (panel (a)) and realized fraction of debt in distress (panel (b)) to model predicted values (realized ■; GLM ♦; GAM ▲; GB ●; GLMM ○). The models are estimated on an expanding window of data with a 2-year gap to the forecasted data set. E.g., the models which are used to forecast the 2011 distresses are estimated on 2003-2009 data. The bars indicate simulated 90% prediction interval where outcomes are simulated using the predicted probabilities for each model.

GB model are virtually the same as the distress rates of the other models in all the other years, suggesting that the GB model is on aggregate similar to the other models. Finally, and perhaps most convincingly, we find no improvements in in-sample results of the GB model compared to the other models in terms of aggregate distress rates. An improvement would be expected in-sample in the case of overfitting.

The amount of debt varies greatly from firm to firm. The largest 21% of the firms have a size greater than 10 million DKK and account for 91% of the total debt in our sample. Thus, the

Table 2.1: Likelihood ratio test for coverage of the out-of-sample 95% quantiles. We form four portfolios of firms representing bank exposures for each calender year yielding 40 portfolios in total. For each portfolio we compute the 95% out-of-sample quantile for the distress rate in each of the three different models and perform a test where the null hypothesis is that the 95% quantiles have the correct coverage level. The "asymptotic p-value" is the p-value from the test in Kupiec (1995) and the "MC p-value" is the Monte Carlo corrected p-values used in Berkowitz et al. (2011).

Model	Likelihood ratio	Asymptotic p-value	MC p-value
GLM	49.670	< 0.0000001	< 0.0000001
GAM	25.901	0.0000004	0.0000004
GB	18.005	0.0000220	0.0000190

percentage of firms in distress and the fraction of debt in distress may be substantially different. Therefore, we also test how the models predict the amount of debt in distress. We compute the fraction of debt in distress each year as

$$\text{DiD}_t = \frac{\sum_{i\in R_t} y_{it}\,(\text{short debt}_{it} + \text{long debt}_{it})}{\sum_{i\in R_t} \text{short debt}_{it} + \text{long debt}_{it}}$$

and the predicted fraction of debt in distress each year for all models as

$$\widehat{\text{DiD}}_{tj} = \frac{\sum_{i\in R_t} \widehat{p}_{itj}\,(\text{short debt}_{it} + \text{long debt}_{it})}{\sum_{i\in R_t} \text{short debt}_{it} + \text{long debt}_{it}}$$

where DiD_t is an abbreviation for "fraction of debt in distress" in year t and $\text{short debt}_{it} + \text{long debt}_{it}$ is the total debt of firm i at time t.

Figure 2.3(b) shows results for the realized and out-of-sample predicted fraction debt in distress. Similarly to the distress level results shown in Figure 2.3(a), we find that none of the models get near the actual level or have 90% prediction intervals with 90% coverage. However, the results here depend highly on a few number of firms. The 25 firms with the largest debt on their balance sheet in 2016 account for 28.47% of the debt. Thus, Figure 2.3(b) essentially reflects a non-trivial probability of default for some of these firms. As seen by Figure 2.3(b), frailty (the GLMM) has little impact with such unequal distributions of exposures. However, we do not expect such unequal distribution of exposures in, say, a bank's loan portfolio. One may suspect that our results are somewhat driven by the latest financial crises. However, Figure 2.3 shows that all three models perform poorly even in the latter part of the sample.

2.5.3 Measuring Portfolio Risk Without Frailty

Above we illustrated that all models fail to capture the percentage of firms entering into distress in the next period. In this section we explore this further by examining the models' ability to evaluate portfolio risks. Specifically, we compare the 95% quantiles of the predicted distress rate distributions to the realized value. If the estimates are accurate, we will find that the realized distress rate is below the upper quantiles about 95% of the cases and above about 5% of the cases.

We use bank connections reported by the firms to construct portfolios for each year and bank. If a firm indicates two bank connections, the firm will appear in the portfolio of both banks. We only include banks with at least 500 connections to ensure that the portfolio is somewhat diversified. Four banks fulfill this requirement. The smallest and largest number of connections for a given bank and year are 534 and 5 063 firms and the mean number of connections is 2 196. We track the four banks through 10 years resulting in a total of 40 portfolios. The portfolios we have constructed are only a rough proxy for the exposure of the banks in the Danish economy. Thus, this exercise should be seen as an example of non-random portfolios rather than as representing the lending risk of the Danish banks.

We define the bank's exposure towards a firm as the reported long-term and short-term bank debt on the firm's balance sheet. A small number of Danish firms issue corporate bonds. The notional of these bonds are included in the bank debt variable in the financial statement, though they are not held by the bank. The notional of the corporate bond is typically much larger than the notional of the actual bank debt, thereby making some firms appear extremely large in the bank debt portfolios. As a simply way of excluding the corporate bonds from the portfolios, we cap the bank debt of the individual firms in each portfolio such that the exposure to a single firm cannot exceed 1% of the total exposure of the bank.

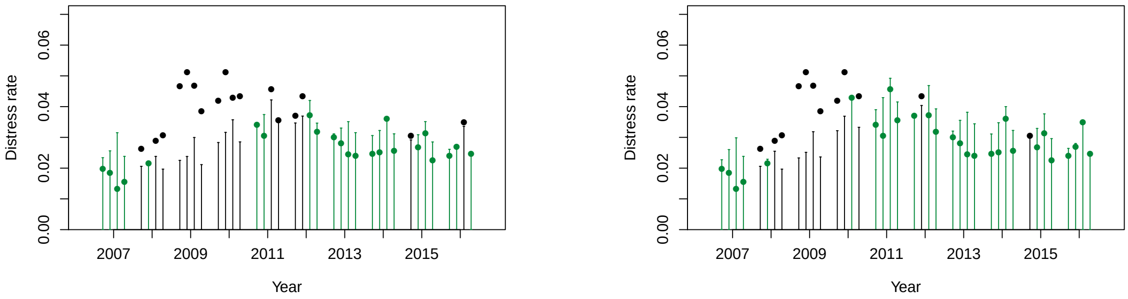
We estimate the out-of-sample 95% quantile of the distress rate in each of the portfolios assuming the GLM, GAM, and GB model respectively and test the coverage of the upper quantiles. Table 2.1 shows results of the coverage test introduced by Kupiec (1995) and the Monte Carlo correction from Berkowitz et al. (2011). We reject the null hypothesis that the coverage has the correct level for all models at a 1% significance level with both the asymptotic p-values and finite sample Monte Carlo corrected p-values. That is, we can statistically reject that any of the models including the GB model are able to estimate accurate risk measures.

Figure 2.4 illustrates when the realized values are above the 95% quantiles for each of the portfolios, where the vertical lines represent 95% quantiles. The lines are green (black) when the realized distress rate is below (above) the upper quantile. The GLM has 17 breaches, the GAM has 12 breaches, and the GB model has 10 breaches. Most breaches occur in 2008-2009.

The models' inability to capture the time-varying distress level and the lack of coverage of the upper quantiles is a sign that the models are misspecified. In order to mitigate this we implement a mixed model in the next section which allows for a random intercept.

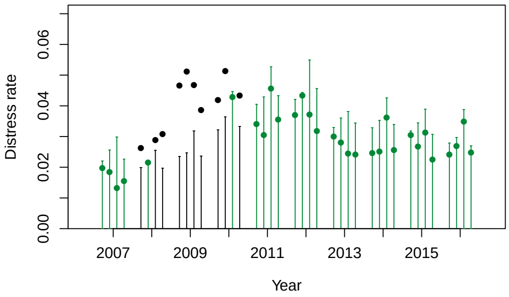
2.6 Modeling Frailty in Distresses with a Generalized Linear Mixed Model

In this section we estimate a generalized linear mixed model (GLMM) introduced in Section 2.3.4 with a random intercept to relax the conditional independence assumption we have assumed so far. That is, the model allows for an unobservable macro effect and thus creates correlations in distresses beyond the observed covariates. Furthermore, we add non-penalized natural cubic regression splines to the model given the higher AUC and lower Akaike information criterion (AIC) of the GAM compared to the GLM (see Appendix 2.B.1 for the latter). While several



(a) 95% quantiles of the distress rate in the GLM

(b) 95% quantiles of the distress rate in the GAM



(c) 95% quantiles of the distress rate in the GB model

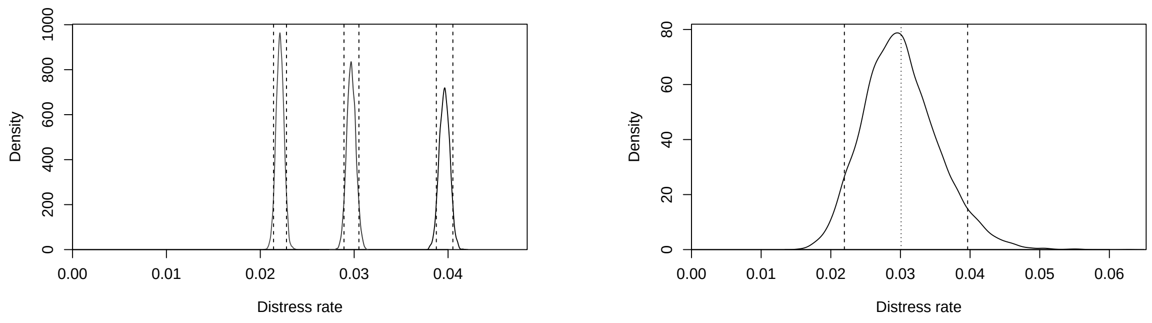
Figure 2.4: Models without frailty estimate too low 95% quantiles. We form bank portfolios based on self-reported bank connections in the firms' financial statements. For each portfolio we compute the 95% quantile by simulation, using the out-of-sample predicted firm probabilities of distress. Panel (a), (b), and (c) show the upper quantiles for the GLM, GAM, and the GB model respectively. The dots show the realized level, bars show the upper quantiles. Black bars and dots indicate years where the realized level is not covered by the prediction interval.

others have implemented GLMM with random intercept or similar random effect models (e.g., see Duffie et al., 2009), we differ by including nonlinear effects. We use non-penalized splines as software allowing for penalized splines in a GLMM is not readily accessible to us and, furthermore, we expect only a minor difference between a penalized and a non-penalized model due to our large sample.

The estimated standard deviation of the random intercept is $\widehat{\sigma} = 0.196$ when estimated on the 2003–2016 data set. That is, a change of one standard deviation in the random intercept implies an $\exp(0.196) = 1.217$ times higher odds of entering into distress for all firms. Thus, there is a non-negligible random effect. A conservative likelihood ratio test for $H_0 : \sigma = 0$ is rejected with a test statistics of 1 483 which is compared to a χ^2 distribution with 1 degree of freedom.[14] Thus, we can reject the conditional independence assumption. We end this section by illustrating what can go wrong if one relies on a model that does not account for the observed correlation in distresses.

---

[14] The p-value is likely conservative (e.g., see the simulations in Pinheiro and Bates, 2000). Though, it does not matter in this case since the p-value is essentially zero already.



(a) Conditional density plots

(b) Unconditional density plots

Figure 2.5: Density plots of the GLMM forecasted 2016 distress rate. We estimate the GLMM on 2003-2016 data and simulate densities of the predicted cross-sectional distress rate in 2016. In panel (a) the random effect is held constant at the 5%, 50%, and 95% quantile of its distribution. The three quantiles can be seen as a "good", "middle", and "bad" future state of the unobservable macro effect in 2016. The tall density curves and narrow prediction intervals are consistent with what a model without a random intercept would predict. Panel (b) shows a density curve estimate where we simulate both the random intercept term and the outcomes. The outer dashed lines are 5% and 95% quantiles and the inner line is the mean.

2.6.1 Predictive Results of the GLMM

Figure 2.5 shows a forecast for the 2016 distress rate and illustrates how adding a random intercept to the model affects the prediction interval of the distress rate. Panel (a) of the figure shows the 2016 forecasts of the distress rate assuming that the random effect is fixed at three different quantiles of its estimated distribution, the 5%, 50%, and 95% quantile. The three quantiles can be seen as a "good", "middle", and "bad" future state of the unobservable macro effect in 2018. Panel (b) of the figure shows the unconditional 2016 forecast density of the distress rate (i.e., without fixing the random intercept). The prediction interval is much wider than that of the GLM, GAM, and GB model (see Figure 2.3(a)), whereas the width of the prediction interval, when the random effect is assumed to take a specific value, is of the same magnitude as in the GLM, GAM, and GB model. The large effect of the random intercept on the prediction interval is similar to what Duffie et al. (2009) find[15] and reflects the large estimated standard deviation of the random effect. It is worth mentioning that there is a large uncertainty in our estimate of the standard deviation of the random intercept, $\widehat{\sigma}$, in the GLMM estimated on the full sample. This is mainly due to the short time series as the GLMM requires a relatively long estimation period and it is not caused by our splines or number of covariates.[16] A 95% profile likelihood-based confidence interval for $\widehat{\sigma}$ is $[0.155, 0.332]$. Typically, accounting for uncertainty in random effect variance yields wider prediction intervals (e.g., see Duffie et al., 2009, p. 2110) so our prediction intervals may be a bit too narrow.

Due to the required estimation period, we can only backtest results of the GLMM in 2016 as in Section 2.5. We will compare these results to results of the GAM in the following, though similar conclusions can be made for the GLM and GB model. In 2016 we find an AUC of 0.815 in the GLMM, which is close to the 0.818 of the GAM we find in 2016. Thus, we find evidence that the GLMM is equally good at ranking the firms in terms of riskiness.

---

[15] See Figure 5 of their paper.

[16] The uncertainty of the estimated standard deviation would be large even if we observed the random effects, ϵ_t.

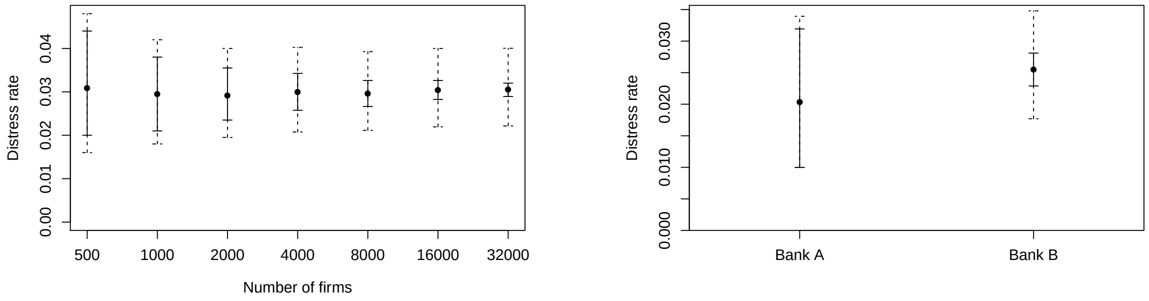
The 90% prediction interval of the distress rate in 2016 is $[0.0220, 0.0396]$ in the GLMM, while we estimated the same interval to be $[0.0302, 0.0317]$ in the GAM. The realized distress rate in 2016 was 0.0318, that is, the realized distress rate is not included in the prediction interval of the GAM while it is included in the prediction interval of the GLMM. Furthermore, the 2016 prediction intervals of the GLMM predicted fraction of debt in distress and the GAM predicted fraction of debt in distress are $[0.007665, 0.02748]$ and $[0.008947, 0.02519]$, respectively, and the realized fraction of debt in distress in 2016 is 0.007113. The prediction intervals of the GLMM and GAM are both illustrated in Figure 2.3. The prediction intervals of the two models are much more similar for the fraction of debt in distress than in the distressed rate. Again, this is due to a few firms in the sample with large debt, implying that a portfolio of firm debt is less diversified. The connection between portfolio diversification and the prediction intervals is explained in the following section.

2.6.2 Frailty Models and Portfolio Risk

Accounting for frailty is more important for some portfolios with distress risk than others. Particularly, adding a frailty to a model matters more for portfolios with many exposures of equal size. To illustrate this point, we randomly sample firms that are active on January 1, 2018 (as defined in Section 2.4.1) into portfolios of sizes ranging from 500 to 32 000. Thus, some portfolios are much more diversified than others, which means that prediction intervals of the predicted distress rate will vary.

For each portfolio we then compute the distress rate using the estimated GLMM and simulate 90% prediction intervals of the distress rate. First, we ignore the frailty component by integrating out the random effect in the firm-specific distress probabilities and draw the firm-outcomes independently using these probabilities. Secondly, a simulation is done where we account for the frailty component by first drawing the random effects from its estimated distribution, compute the firm probabilities conditional on the drawn random effect, and then draw the firm outcomes conditional on these probabilities. The second method is the same as the one used for the simulated prediction intervals in Figure 2.5(b), and the width of the prediction intervals of the model without frailty is very similar to the width of the prediction intervals in Figure 2.5(a), which again is very similar to the prediction intervals of the GB model.

The results are shown in Figure 2.6(a). The figure illustrates that the tail risk is generally underestimated when we do not account for frailty. However, the discrepancy between the two models is much more pronounced for the large portfolios than for the small. This is because the model without frailty drastically shrinks the prediction intervals of the more diversified large portfolios. The prediction intervals of the model with frailty are also affected when the portfolio becomes more diversified, but to a much smaller extent. This is because the frailty model accounts for the excess clustering defaults because of the latent variable. An economist relying on a model without frailty could then easily conclude that a well diversified portfolio is much safer than what it is in reality. How this can lead to misperception of portfolio risk of two banks with different



(a) Random portfolio example **(b)** Example with reversed ranking

Figure 2.6: Frailty matters more for large portfolios. In panel (a), we randomly split the sample of firms that are active on January 1, 2016 into portfolios of size $500, 2 \cdot 500, \dots, 32\,000$ and compute the distress probability based on the GLMM estimated on the 2003-2016 sample. The dots are the expected unconditional distress rate of the portfolios. The solid lines are the simulated 90% prediction interval where we integrate the random effect out on a firm-by-firm level and then simulate the outcomes independently. The dashed lines are the simulated prediction interval when we do account for frailty. Panel (b) shows 90% prediction intervals of the distress risk of loan portfolios of two banks. Bank A has 501 clients with distress probabilities evenly distributed on the interval $[0.10, 0.30]$ on the logit scale in the case of the GLMM where the random effect is equal to zero. Likewise, Bank B has 10 001 clients with distress probabilities evenly distributed on the interval $[0.15, 0.35]$ when the random effect is zero. The solid and dashed lines are prediction intervals simulated in a model without and with frailty respectively.

strategies is illustrated in the following example.

Assume that we have two banks: Bank A has a few safe clients and Bank B has many relatively more risky clients. Specifically, Bank A has provided a loan to 501 clients with distress probabilities evenly distributed on the interval $[0.10, 0.30]$ in the case of the GLMM where the random effect is equal to zero. Bank B has provided a loan to 10 001 clients with distress probabilities evenly distributed on the interval $[0.15, 0.35]$ when the random effect is zero. Appendix 2.C provides further details regarding the simulation of the two bank portfolios. The prediction intervals of the distress rate with and without accounting for frailty are illustrated in Figure 2.6(b). The 95% quantiles for Bank A and Bank B are 0.0319 and 0.0281 respectively, if we do not account for frailty. Thus, Bank A appears more risky by this metric. However, the correct figures – the ones where we account for frailty – are 0.0339 and 0.0348 respectively. Hence, Bank B has the highest risk by this metric in reality. Thus, if one relies on a model without frailty, one might wrongly assume that a large bank is exposed to relatively little risk.

2.7 Including Macro Variables in the Models

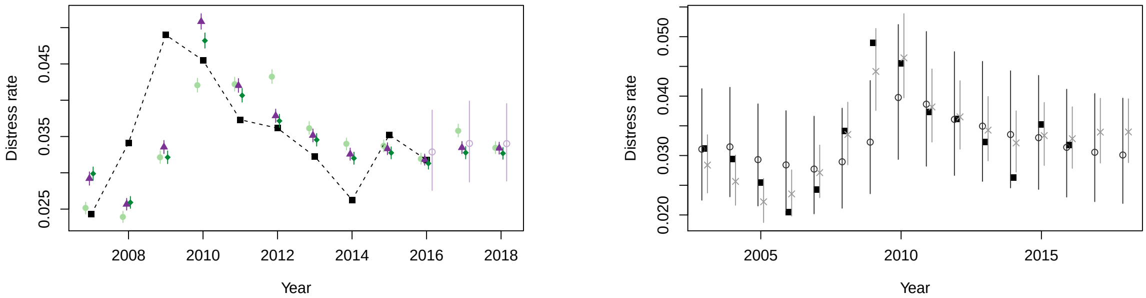
The models implemented so far are based solely on micro level data. While the models are good at ranking the individual firms by riskiness, they are far from good at estimating the aggregated distress rate in the next period. This raises the question whether the models could be improved by including some macro variables. In this section we show results of models including macro variables, and we argue that we may have a potential candidate for macro variable but estimating an effect of the covariate may be hard with the few number of cross-sections we have. Lastly, we show that the random effect is still needed in-sample after the inclusion of the random effect.

Some common macro variables in the existing literature are return of the S&P 500 index, 3-month treasury rate, 10-year treasury rate, inflation, GDP growth, and the unemployment rate (e.g., see Chava et al., 2011, Duan et al., 2012, Duffie et al., 2009, Filipe et al., 2016, Lando et al., 2013). We include the Danish equivalent of these variables in our models, except for the stock index return since the majority of firms in our sample are non-traded, and test if the models' predictions improve. We lag all macro variables to ensure predictability. Furthermore, we use a swap rate, a short-term, and a long-term interbank rate instead of treasury rates, since the Danish government bond market is much smaller than the U.S. Finally, we include the GDP gap instead of the GDP growth as GDP gap has been included in earlier versions of the Danish central bank's internal corporate distress model. While some of the aforementioned papers track events on a quarterly or monthly basis, we choose to do so only on a yearly basis. This is because the start date of the "in distress" status can be somewhat arbitrary and reflects a potentially delayed processing time of the authorities. Thus, we end up with relatively few observations in the time dimension, implying that we can include at most one macro variable in our models.

We run separate logistic regressions including each of the macro variables one at a time and find that the model with the unemployment rate has the lowest AIC. We then include the unemployment rate in all four models and run predictive tests. The inclusion of the unemployment rate in the GB model is done by estimating a logistic regression with two covariates: the unemployment rate and the linear predictor from the estimated GB model without the unemployment rate. The motivation for the two-step model is that we can control the complexity of the unemployment rate. This turned out to be an issue in some preliminary results where a GB model including the unemployment rate as a covariate generalized poorly. Including a macro variable in the GB model could potentially lead to a separate model for each year which likely would generalize poorly.

We find improvements in the out-of-sample forecasts when the unemployment rate is included in the models (see Figure 2.3(a) and Figure 2.7(a)). However, Figure 2.7(a) still shows too narrow prediction bounds for the GLM, GAM and GB model. The standard deviation of the random intercept of the GLMM estimated in the period 2003 to 2016 is reduced from 0.196 in a model without the unemployment rate to 0.106 in a model with the unemployment rate. The reduction shows that the unemployment rate explains some of the yearly fluctuations. This is also evident from Figure 2.7(b) which shows a much better in-sample predicted distress rate for the GLMM with the unemployment rate. However, the random intercept remains significant with a χ^2 test-statistic of 297 with 1 degree of freedom.

The estimated slope on the unemployment rate is negative and statistically significant, which may seem counter-intuitive. Furthermore, the slope estimate varies a lot during the first out-of-sample forecasts, which is not surprising given the low number of cross-sections included in this sample. One major question is whether we will see the same in the future, i.e., if the association we estimate now will generalize. This is particularly questionable given that we have already considered five potential macro variables with only 14 cross-sections. However, we also estimate



(a) Realized and predicted distress rate **(b)** In-sample predicted distress rate

Figure 2.7: Results of models with the unemployment rate. We re-estimate all four models adding the unemployment rate as a covariate. Panel (a) shows the realized distress rate together with the out-of-sample predicted values (realized ■; GLM ♦; GAM ▲; GB ●; GLMM ○). The bars indicate the simulated 90% prediction interval where outcomes are simulated using the predicted probabilities for each model respectively. Panel (b) shows the predicted distress with and without the unemployment rate along 90% prediction intervals for the GLMM with parameters estimated on the full sample (realized ■; GLMM without unemployment rate ○; GLMM with unemployment rate ×).

a negative slope for the unemployment of the same size on aggregate defaults for which we have data going back to 1980.[17] This provides evidence that the effect we estimate may generalize. Whether the estimated slope on unemployment generalizes or not does not change the fact that the random intercept remains significant, i.e., we cannot avoid a frailty component.

2.8 Conclusion

We have shown that gradient tree boosting performs better in out-of-sample ranking of firms in terms of riskiness compared to more traditional statistical models in a sample containing the majority of Danish limited liability firms. However, the improvement is only minor compared to what recent papers find. Furthermore, the out-of-sample tests yield too narrow prediction intervals of the aggregated distress rate for both traditional statistical models and the gradient boosted tree model. That is, the more complex model is not better at capturing correlations in defaults across the cross-section of firms, leading to too small risk measures for individuals, firms, or regulators who evaluate the riskiness of a portfolio exposed to multiple firms. We show how to relax this assumption with a generalized linear mixed model, where we relax the linearity assumption for some the covariates in the model, thereby obtaining competitive firm-level performance.

While Basel III does incorporate correlations in defaults, it is not obvious that the same correlations should be used regardless of the model that is used to produce the marginal probabilities of defaults. As an example, Lando and Nielsen (2010) fail to reject the miss-specification tests in Das et al. (2007) after inclusion of additional covariates. That is, the level of excess clustering of defaults may vary depending on the covariates in the model. Similar effects can occur with time-varying covariate distributions and a partial effect which is erroneously assumed to be a line or a plane. The final frailty model in this paper is an example of a model which can relax the linearity assumption and adjust the excess clustering of defaults depending on the assumed association between covariates and the log odds of a distress event.

---

[17] We use the number of VAT registered firms (Danmarks Statistik, 2018a) as the denominator and the number of defaults (Danmarks Statistik, 2018b) as the numerator in a binomial regression model similarly with the logit link function.

Appendix

2.A Variable Selection with Lasso

We use the so-called Thresholded Lasso estimator to perform variable selection. The Thresholded Lasso estimator is found by the following steps.

1. Standardize the covariates.
2. Perform K-fold cross-validation to find the penalty variable, λ_{init}, that maximizes

$$\lambda_{\text{init}} = \arg\max_{\lambda} \max_{\boldsymbol{\beta}} \sum_{t=1}^{d} \sum_{i \in R_t} y_{it} \boldsymbol{\beta}^\top \boldsymbol{x}_{it} - \log\Big(1 + \exp\Big(\boldsymbol{\beta}^\top \boldsymbol{x}_{it}\Big)\Big) - \lambda \|\boldsymbol{\beta}\|_1$$

   where $\|\cdot\|_1$ is the L1 norm. This is the log-likelihood from the multiperiod logit model in Equation (2.1) with an added L1 penalty.

3. Denote $\widehat{\boldsymbol{\beta}}_{\text{init}}$ as the estimated coefficients with penalty parameter λ_{init} and define the sets $S(\delta) = \{j : |\beta_{\text{init},j}| > \delta\}$. Then use K-fold cross-validation to find a threshold value, $\widehat{\delta}$, in the range that maximizes

$$\arg\max_{\delta} \max_{\boldsymbol{\beta}_{S(\delta)}} \sum_{t=1}^{d} \sum_{i \in R_t} y_{it} \boldsymbol{\beta}_{S(\delta)}^\top \boldsymbol{x}_{S(\delta),it} - \log\Big(1 + \exp\Big(\boldsymbol{\beta}_{S(\delta)}^\top \boldsymbol{x}_{S(\delta),it}\Big)\Big)$$

   where $\boldsymbol{\beta}_{S(\delta)}^\top \boldsymbol{x}_{S(\delta),it}$ is the linear predictors which only include the covariates in the index set $S(\delta)$. This amounts to fitting a GLM with a subset of the covariates.

Step 1 and 2 yield the common Lasso estimator, $\widehat{\boldsymbol{\beta}}_{\text{init}}$. That is, we add an L1 penalty which shrinks parameters and discards variables where the coefficient is shrunk to 0. Step 3, in addition to the previous, yields the Thresholded Lasso estimator, where we discard any variables where the coefficient is below the threshold $\hat{\delta}$. The final estimates are no longer shrunk as we do not apply a penalty. The motivation to use the Thresholded Lasso estimator rather then the Lasso estimator is to addresses the bias problems with $\widehat{\boldsymbol{\beta}}_{\text{init}}$. See Bühlmann and Van De Geer (2011), Zhou (2010) for properties of the Thresholded Lasso estimator.

We end with the 6 categorical and 44 numerical covariates listed in Table 2.2. The numerical variables are divided by firm size when appropriate and all are winsorized at the 5% and 95% quantile. We exclude 3 numeric covariates in the Lasso estimation while none of the categorical covariates considered are dropped.

Table 2.2: Summary statistics for covariates in the data set from 2003 to 2016. Variables divided by size are in percentages. Size is the maximum of total asset and total debt. The statistics are computed after winsorizing. There is 1.3 million firm year observations. Panel A shows the numerical covariates that are left after variable selection with the Thresholded Lasso method and the estimated coefficients where stars indicate the significance of the effect with a Wald test (*** is 1% significance, ** is 5%, and * is 10%). Panel B shows the numerical variables that are excluded after variable selection. Panel C shows the binary and categorical covariates included in the model. Panel D shows variable descriptions of some of the covariates.

Covariate	Mean	Median	St. Dev.	Min	Max	GLM coefficient estimates
Panel A: Numerical covariates included after variable selection						
Accounts payable / size	8.06	2.83	10.99	0.00	38.00	0.0246***
Accounts receivable / size	12.75	3.34	16.88	0.00	54.00	−0.0097***
Change in log size	0.03	0.00	0.24	-0.46	0.58	−0.0901**
Corporation tax / size	1.12	0.00	2.18	0.00	7.60	0.0708***
Current assets / size	58.41	66.07	35.50	1.00	100.00	−0.0025***
Deferred tax / size	1.16	0.00	2.31	0.00	8.10	−0.0688***
Depreciation / size	-3.08	-1.11	4.18	-14.00	0.00	0.0140***
EBIT / size	4.19	3.44	17.44	-36.00	40.00	−0.0036***
Equity / invested capital	6.16	2.27	10.05	-3.90	38.00	−0.0255***
Equity / size	33.78	32.32	38.10	-48.00	96.00	0.0032***
Expected dividends / size	1.59	0.00	4.08	0.00	15.60	−0.0968***
Financial assets / size	6.10	0.00	14.71	0.00	58.00	−0.0117***
Financial income / size	0.99	0.17	1.63	0.00	5.80	−0.0531***
Financing costs / size	2.22	1.54	2.25	0.00	7.40	0.0479***
Fixed costs / size	-44.96	-25.09	52.16	-175.00	0.00	0.0002
Immaterial fixed assets / size	1.73	0.00	4.82	0.00	19.00	−0.0169***
Ind. EW avg. net profit / size	2.03	2.11	2.94	-39.00	34.00	−0.0279***
Interest coverage ratio	0.02	-0.71	21.75	-47.00	48.00	−0.0003
Inventory / size	9.09	0.00	16.45	0.00	56.00	−0.0052***
Invested capital / size	20.16	9.40	25.77	0.90	97.00	0.0009**
Land and buildings / size	16.04	0.00	31.17	0.00	95.00	−0.0076***
Liquid assets / size	14.94	3.51	21.69	0.00	75.00	−0.0131***
log(age)	1.98	2.08	1.16	0.00	4.60	−0.2965***
log(size)	7.85	7.82	1.61	4.95	10.91	0.0133*
Long-term bank debt / size	2.60	0.00	6.98	0.00	26.00	0.0110***
Long-term debt / size	11.65	0.00	20.49	0.00	66.00	0.0026***
Long-term mortgage debt / size	5.20	0.00	13.24	0.00	47.00	0.0077***
Net profit / size	2.04	2.16	16.52	-39.00	34.00	−0.0066***
Other operating expenses / size	-2.24	0.00	6.08	-23.00	0.00	−0.0094***
Other receivables / size	4.33	0.97	7.15	0.00	26.00	−0.0050***
Other short debts / size	13.79	8.58	15.01	0.00	53.00	0.0114***
Personnel costs / size	-34.28	-10.05	45.82	-151.00	0.00	0.0015***
Prepayments / size	0.52	0.00	0.96	0.00	3.40	−0.0659***
Provisions / size	1.34	0.00	2.60	0.00	9.20	0.0096*
Quick ratio	2.35	0.98	3.86	0.00	16.00	−0.0040
Receivables from related parties / size	5.61	0.00	12.99	0.00	49.00	0.0014**
Relative debt change	0.10	0.00	0.48	-0.62	1.50	−0.0621**
Retained earnings / size	6.20	6.73	38.71	-91.00	72.00	−0.0049***
Return on equity (pct.)	-1.05	0.12	4.95	-19.40	3.60	−0.0175***
Short-term bank debt / size	7.32	0.00	13.19	0.00	44.00	0.0109***
Short-term mortgage debt / size	0.12	0.00	0.38	0.00	1.50	−0.0223
Tangible fixed assets / size	26.17	9.44	32.59	0.00	96.00	−0.0040***
Tax expenses / size	-1.68	-0.44	3.51	-10.30	3.80	−0.0080***
Total receivables / size	26.91	18.69	26.82	0.00	90.00	0.0037***

Continued on next page

Table 2.2 – *Continued from previous page*

Covariate	Mean	Median	St. Dev.	Min	Max	GLM coefficient estimates
Panel B: Numerical covariates excluded after variable selection						
Current ratio	2.58	1.21	3.87	0.00	16.00	
max(equity + provisions, 0) / size	39.74	34.79	31.88	0.00	100.00	
Short-term debt / size	47.94	45.87	32.11	1.80	100.00	
Panel C: Categorical covariates						
Is non-stock based	0.73	1.00	0.45			0.3277***
Has prior distress	0.03	0.00	0.16			0.9542***
Large debt change	0.08	0.00	0.27			0.1936***
Negative equity	0.15	0.00	0.36			0.1770***
Region						
Sector						

Panel D: Variable description

Change in log size	The log of firm size as reported in the current financial account minus the log of firm size as reported in the financial account from the previous year. We use the size definition in Equation (2.8). The variable is set to zero if the firm did not hand in a financial account the previous year.
Current ratio	Current assets divided with short-term debt. If the short-term debt is zero or below 10 000 DKK we divide by 10 000 instead to avoid dividing with zero.
Is non-stock based	A dummy variable equal to 1 if the firm is non-stock based (“Anpartsselskab”). The alternative is a stock-based firm (“Aktieselskab”).
Has prior distress	A dummy variable equal to 1 if the firm has previously been “in distress”.
Ind. EW avg. net profit	We group firms by their 3-digit SIC code and compute the equally weighted average net profit of each group each year.
Interest coverage ratio	Net profit divided by net financial revenue. If the net financial revenue is zero, we divide by 1 instead.
Large debt change	A dummy variable equal to 1 if the total debt grew more than 100% in the past year. It is zero if the firm did not hand in a financial account the previous year.
Negative equity	A dummy variable equal to 1 if equity is negative.
Region	The firms are grouped based on the location of their headquarter into 5 geographical regions, going from the most to the least densely populated areas.
Relative debt change	The firm’s total debt of the current financial account divided by the total debt of the financial account from the previous year. The variable is set to zero if the firm did not hand in a financial account the previous year.
Return on equity	Net profit divided by equity. If equity is zero or below 10 000 DKK we divide by 10 000 instead to avoid dividing with zero.
Sector	The firms are grouped into 7 general sectors: Construction; industrial; farming and fishing; trade; transport; information; real estate; other.
Quick ratio	Current assets minus inventories divided with short-term debt. If the short-term debt is zero or below 10 000 DKK we divide by 10 000 instead to avoid dividing with zero.

2.B Model Estimation

The estimated coefficients of the covariates included in the GLM after variable selection are listed in Table 2.2. Figure 2.8 shows the largest absolute standardized coefficients. Most noticeably, we find a large age effect unlike Shumway (2001). This is not surprising given that we use the age since incorporation for some potentially small and risky firms whereas Shumway (2001) uses the age since listing for large corporate firms but mentions that the results may differ if age since incorporation was used. An age effect is also found in Filipe et al. (2016) who model distress events of SMEs. The industry specific covariate mentioned in Section 2.4.2 has a coefficient estimate of -0.02791 with a standard error of 0.00176. The negative sign is consistent with the results in Chava et al. (2011) who find a higher likelihood of a default when the median stock performance in an industry is below 20%.

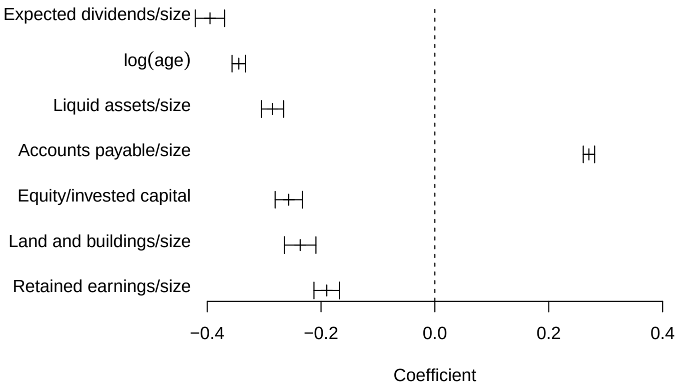


Figure 2.8: Standardized coefficients in the GLM. The plot shows the effect on the linear predictor from a one standard deviation move in the covariates. Only the 7 largest absolute standardized estimates of non-dummy variables are included in the plot. The outer lines are 95% Wald confidence intervals and the inner lines are the estimated coefficients.

The GLM uses 61 degrees of freedom whereas the GAM uses 254.7. Hence, the GAM is much more complex than the GLM.[18] In-sample estimations on the full sample period (2003-2016) yield Akaike information criterion (AIC) of the GLM and GAM at 327 625 and 321 388 , respectively, which shows an improvement from the GLM to the GAM in spite of the increased complexity of the model.

2.B.1 Nonlinear Effects in the GAM and the GLMM

We will show the estimated nonlinear effects in GAM and GLMM in this section. The main objective is to show that we can easily draw inference about the partial effects. This is unlike what we can do for the GB model with trees with a maximum depth of three or deeper. All partial effects of nonlinear terms will be shown. The GAM we estimate can be decomposed into the following ANOVA decomposition

$$\eta_{it} = \boldsymbol{\beta}^\top \boldsymbol{x}_{it}^{(f)} + \sum_{j=1}^{q} \boldsymbol{\gamma}_j^\top \boldsymbol{f}_j(x_{itj}^{(s)}) + \sum_{(j,j_1,j_2)\in J} \boldsymbol{\omega}_j^\top \operatorname{vec}\left(\boldsymbol{f}_{j_1}(x_{itj_1}^{(s)}) \otimes \boldsymbol{f}_{j_2}(x_{itj_2}^{(s)})\right) \tag{2.10}$$

where vec(·) is the vectorization operator, J is the set of covariate pairs with a tensor product spline, and $\boldsymbol{\beta}$, $\boldsymbol{\gamma}_j$, and $\boldsymbol{\omega}_j$ are unknown parameters. This is a generalization of the linear predictor

---

[18] We use the effective degrees freedom, which is $\operatorname{tr}(\mathbf{F}_{\boldsymbol{\lambda}})$ from Equation (2.4) for the GAM.

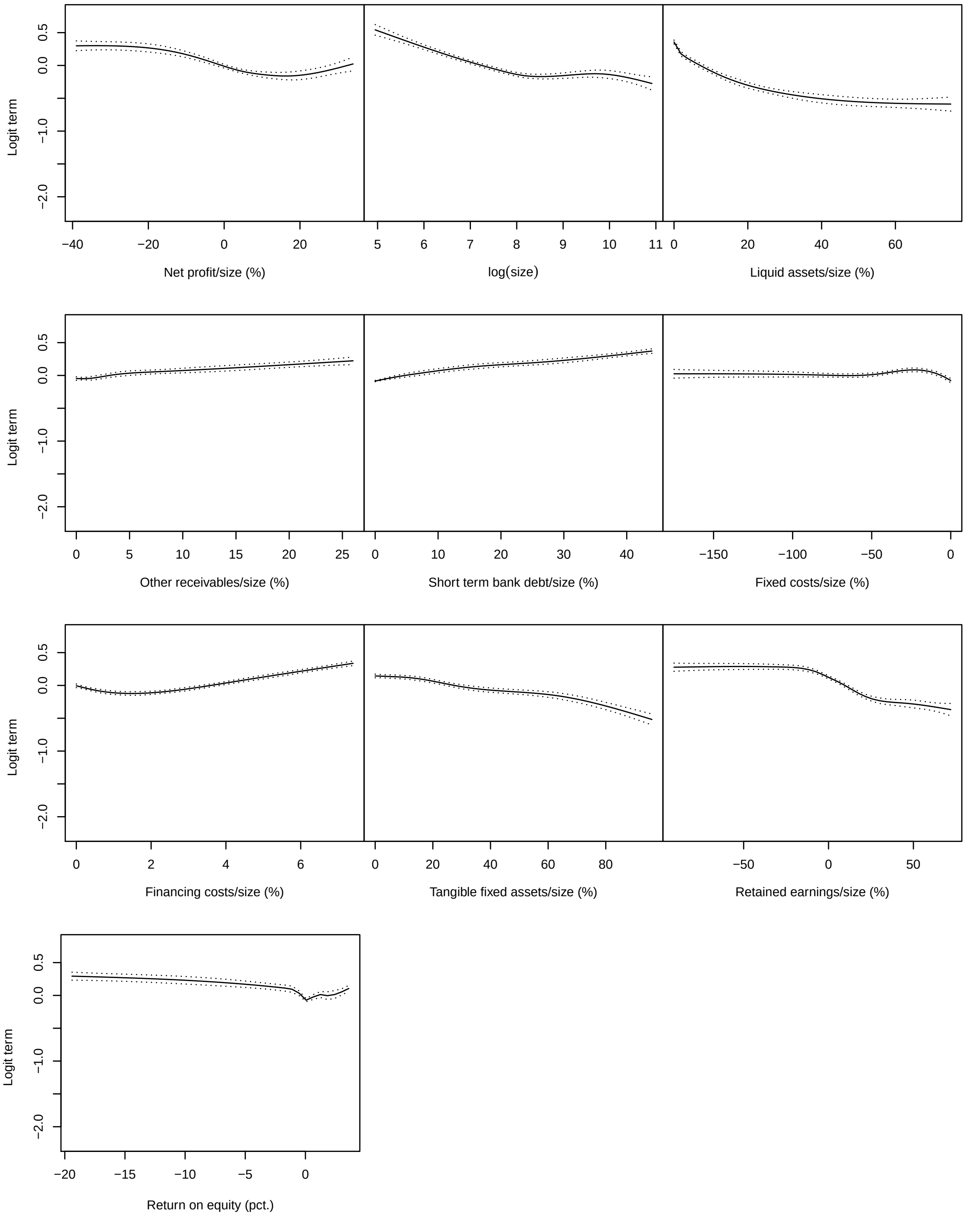


Figure 2.9: Main effects in GAM. The plots show the estimated main effects on the log odds scale of the probability of entering into distress as in Equation (2.10). The fully drawn line is estimated main effect and the dashed lines are ±2 standard deviations conditional on the estimated penalty variables in $\boldsymbol{\lambda}$. The spline bases are subject to a sum-to-zero constraint.

in Equation (2.3) to handle tensor product splines. Each basis function, $\boldsymbol{f}_j$, is subject to a sum-to-zero constraint which is a generalization of centering in the univariate cases. Thus, the $\boldsymbol{\gamma}_j^\top \boldsymbol{f}_j(x_{itj}^{(s)})$-terms can be interpreted as main effects and the $\boldsymbol{\omega}_j^\top \operatorname{vec}\left(\boldsymbol{f}_{j_1}(x_{itj_1}^{(s)}) \otimes \boldsymbol{f}_{j_2}(x_{itj_2}^{(s)})\right)$-terms can be interpreted as interaction effects similarly as in the linear case with centered covariates. A

Table 2.3: Nonlinear effects in the GAM and the GLMM. All covariates except for log(size), log(age), and return on equity are divided by the size variable defined in Equation (2.8). "Varying coefficient" implies an interaction with a spline basis function for the first covariate and the identity function for the second covariate.

First covariate	Second covariate	Type of term
Nonlinear effects in the GAM		
Retained profit		Spline
Return on equity		Spline
Net profit	log(age)	Varying coefficient
Net profit	log(size)	Tensor product spline
Net profit	Liquid assets	Tensor product spline
Net profit	Other receivables	Tensor product spline
log(size)	Short-term bank debt	Tensor product spline
log(size)	Financial costs	Tensor product spline
log(size)	Tangible fixed assets	Tensor product spline
Liquid assets	log(size)	Tensor product spline
Liquid assets	Fixed costs	Tensor product spline
Nonlinear effects in the GLMM		
Retained profit		Spline
Return on equity		Spline
Liquid assets		Spline
Other receivables		Spline
Financial costs		Spline
Net profit	log(size)	Tensor product spline
log(size)	Short-term bank debt	Tensor product spline

so-called varying coefficient is where one of the basis functions, $\boldsymbol{f}_j$, is the identity function. Thus, the slope of the covariate with the identity function as the basis function can be seen as varying as a function of the other covariate.

The GAM has 11 nonlinear effects of which nine have interactions. Table 2.3 shows all the nonlinear effects. We reject tests as suggested by Wood (2013) at the conventional 5% level and the largest p-value is less than 10^{-8}. The null hypothesis is that the true effect is a line or a plane. Figure 2.9 shows the estimated main effects. Multiple of the main effects show an expected monotone effect, e.g., the log size plot shows a decreasing partial association between distress and the size of the firm. Though, the effect is clearly nonlinear for some of the main effects contrary to the assumption in the GLM. The net profit to size ratio is an example that has a slightly non-monotone estimated effect. There is a strong association between the distress rate and changes in the net profit ratio in the range from -10% to 0%, while the dependence of the linear predictor is flat in other regions on the net profit ratio scale. Also, firms with a very high net profit ratio tend to have a slightly higher rate of distress. A potential explanation is that firms with relatively large profits may be more volatile firms.

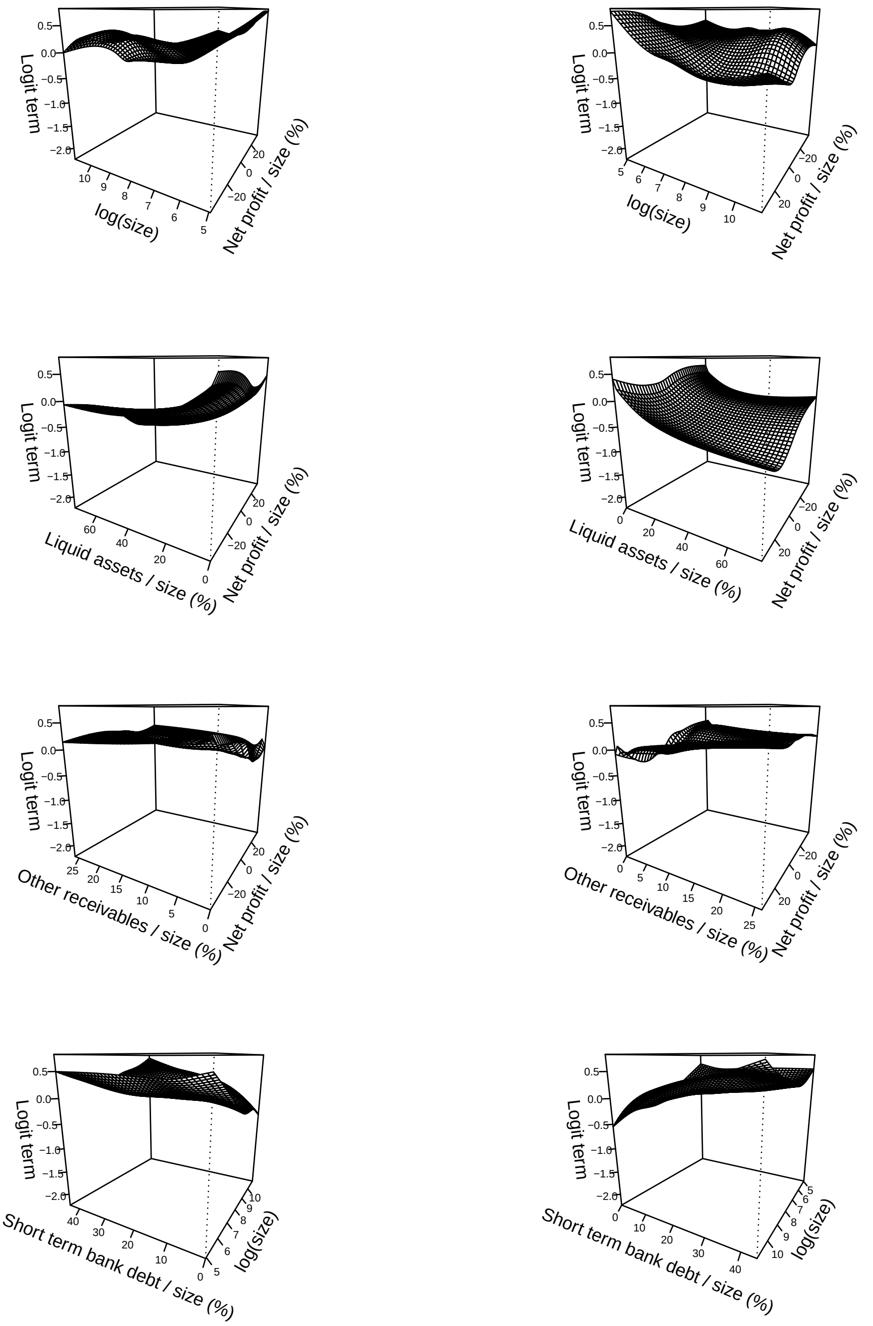


Figure 2.10: Main and interaction effects in GAM. The plots show the estimated main effects plus interaction on the log odds scale of the probability of entering into distress as in Equation (2.10). Confidence intervals as in Figure 2.9 are omitted but could easily be computed in a similar manner. The z-axes are similar to the y-axes in Figure 2.9 for comparisons. The plots on the right side are the same tensor product splines rotated 180 degrees.

Figure 2.10-2.12 shows the main and interaction effects. That is, each plot shows

$$\boldsymbol{\omega}_j^\top \operatorname{vec}\left(\boldsymbol{f}_{j_1}(x_{itj_1}^{(s)}) \otimes \boldsymbol{f}_{j_2}(x_{itj_2}^{(s)})\right) + \boldsymbol{\gamma}_{j_1}^\top \boldsymbol{f}_{j_1}(x_{itj_1}^{(s)}) + \boldsymbol{\gamma}_{j_2}^\top \boldsymbol{f}_{j_2}(x_{itj_2}^{(s)})$$

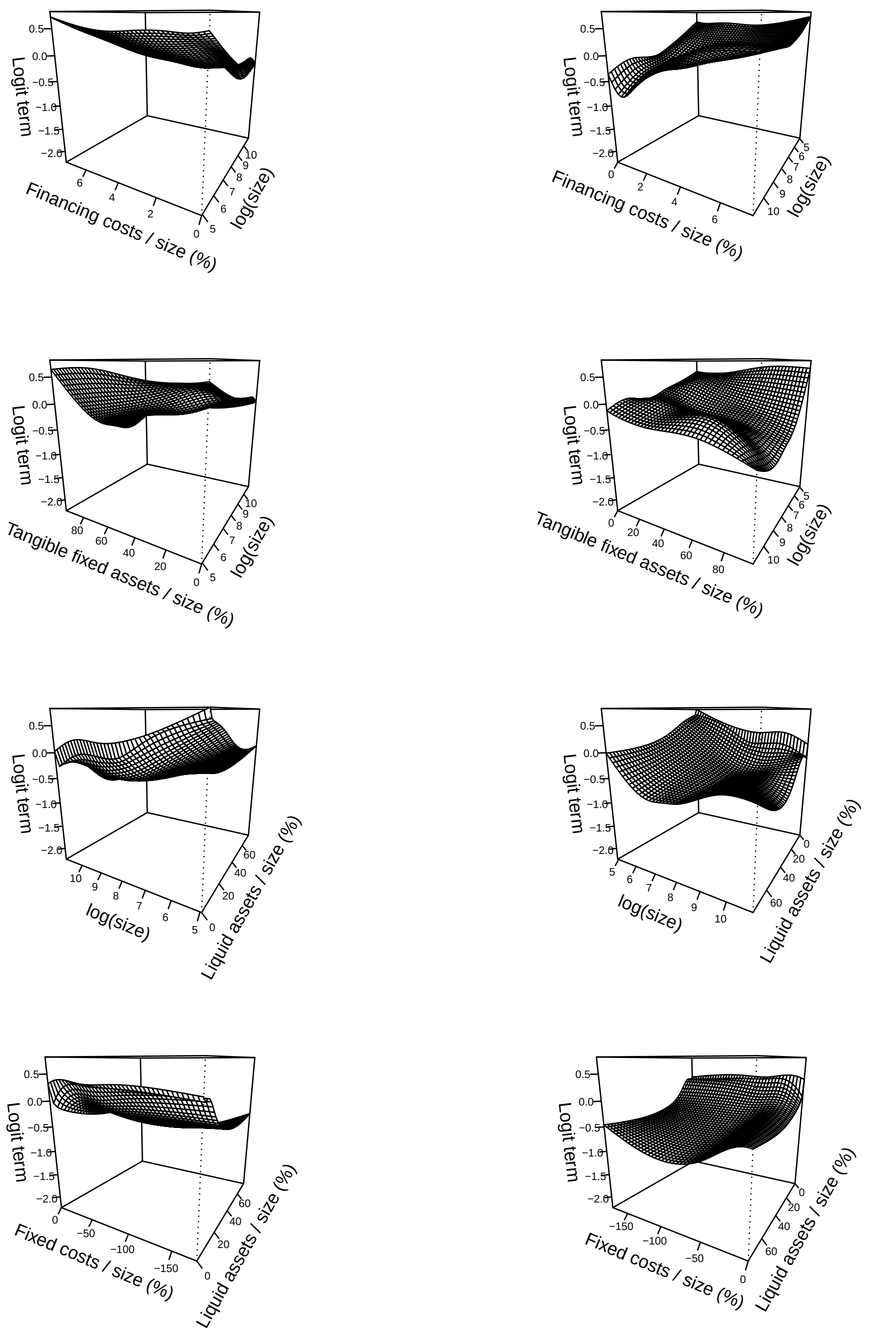


Figure 2.11: Main and interaction effects in GAM. Similar plots as in Figure 2.10.

for given $(j, j_1, j_2) \in J$. There is clear evidence of an interaction effect between some variables. E.g., consider the log size and net profit ratio in Figure 2.10. Firms with large relative losses have almost no partial association between the size of the firm and the probability of entering into distress.

We stress that parts of the covariate space is very sparse in the sense of having a low number of observations or a low number of observed events in the 3D plots shown in Figure 2.10-2.12. Thus, there is some uncertainty about the estimated partial effects in these areas. Finally, we

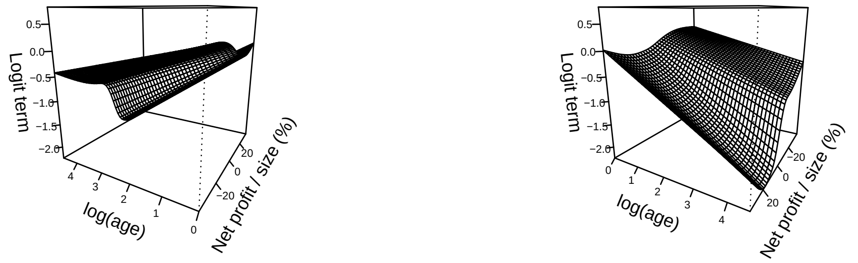


Figure 2.12: Main and interaction effects in GAM. Similar plots as in Figure 2.10.

include 6 nonlinear effects in the GLMM model with only 2 nonlinear interactions. The included effects are shown in Table 2.3. We use 6 to 20 dimensional basis for each spline basis, $\boldsymbol{f}_j$, in the GAM while we only use 5 in the GLMM. The splines in the GAM are selected by inspection of standard diagnostic plots.

We have reduced the dimension of each spline in the GLMM compared to the GAM as we do not penalize the splines in the GLMM. The splines and tensor product splines we include in the GLMM are based on visual inspections of the estimated effect in the GAM and an estimate of the marginal and joint density of pairs of covariates. The estimated splines in the GLMM are comparable to the GAM. The software we use to estimate the GLMM is single-threaded and does not readily have methods to impose penalties as the software we use to estimate the GAM. One could implement a multi-threaded version of the GLMM estimation method and potentially use approximations to reduce the computation time, however, this is beyond the scope of this paper.

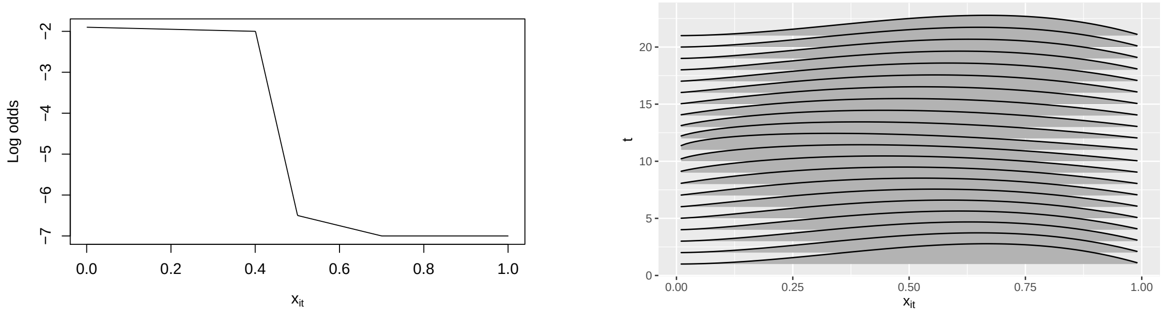
2.C Details of the Two Bank Portfolios Example of Section 2.6.2

We will provide details regarding the simulation example in Section 2.6.2 in the following. Assume that we have two banks: one with few low-risk loans (Bank A) and one with many high-risk loans (Bank B). The probability of a default for each firm j in bank portfolio i is

$$p_{ij} = g^{-1}(\eta_{ij} + \epsilon), \qquad \epsilon \sim N(0, \sigma^2), \qquad \eta_{ij} = g\left(\frac{\bar{p}_i(j-1) + \underline{p}_i(n_i - j)}{n_i - 1}\right)$$

where ϵ is a random effect which we cannot observe and g is the logit function. We fix σ to 0.2 and let Bank A have $n_A = 501$ clients and Bank B have $n_B = 10\,001$ clients. Furthermore, we set Bank A's risk parameters to $(\underline{p}_A, \bar{p}_A) = (0.10, 0.30)$ and Bank B's risk parameters to $(\underline{p}_B, \bar{p}_B) = (0.15, 0.35)$. Thus, the latter bank has more clients which are more risky on average. We use the marginal firm probabilities when we simulate the firms' outcome in the model that does not account for frailty. These probabilities are given by

$$\tilde{p}_{ij} = \int_{\mathbb{R}} g^{-1}(\eta_{ij} + \epsilon)\varphi(\epsilon; \sigma^2) d\epsilon > g^{-1}(\eta_{ij})$$



(a) Log odds conditional on X_{it}

(b) Density of X_{it}

Figure 2.13: Log odds and covariate distribution. Plot (a) shows the log odds as defined in Equation (2.11). Plot (b) shows the density of X_{it} for $t = 1, \dots, 21$.

where $\varphi(\cdot; \sigma^2)$ is the normal distribution density function with zero mean and variance σ^2 and the in-equality follows from a Jensen's inequality and holds when $\eta_{ij} < 0$ (i.e., the probability of default is less than 50%). The firms' outcome are then simulated independently using $\tilde{p}_{ij}$ to produce prediction interval for the portfolios similar to what we do for the GLM, GAM, and GB model in e.g., Figure 2.3.

2.D Example of Nonlinear Effect

One argument in this article is that a misspecification due to e.g., a linearity assumption can yield biased results which may lead one to incorrectly conclude that there is evidence of a frailty or a macro effect. We provide an example in this section to illustrate that this is possible. This is somewhat different from Lando and Nielsen (2010) who show that omission of covariates may yield evidence of a macro effect or a frailty variable.

Suppose we have a single covariate $X_{it} \in (0, 1)$ and that

$$\begin{aligned} z_+ &= \max(z, 0) \\ \operatorname{logit} \mathrm{P}\left(Y_{it} = 1 \mid X_{it} = x\right) &= -1.9 - 0.25x - 44.75(x - 0.4)_+ + 42.5(x - 0.5)_+ + 2.5(x - 0.7)_+ \end{aligned} \tag{2.11}$$

Further, we assume that covariates, $X_{it}s$, are independent and distributed according to

$$\begin{aligned} X_{it} &\sim \mathrm{Beta}\left(\alpha(t), 2\right) \\ \alpha(t) &= 1.5 \left| \frac{t - 11}{10} \right| + 1.5 \end{aligned}$$

where $t = 1, \dots, 21$. Figure 2.13 shows the log odds in Equation (2.11) and the density of X_{it} for $t = 1, \dots, 21$. We assume that events are recurrent so an individual is at risk after having an event. Given this model we make 1000 simulation with 10000 individuals in each sample ($i = 1, \dots, 10000$). We fit two models in each simulation: one which assumes a linear effect of X_{it} on the log odds scale and one where we add a second order polynomial for time. The latter can be seen as a macro variable or a frailty effect. Lastly, we perform a likelihood ratio test between the two models. We reject the test in 869 of the 1000 simulation at the conventional 5 pct. level which could, incorrectly, be inferred as a macro or a frailty effect.

While the example may be simple, it does show that an incorrectly specified partial association combined with a time-varying covariate distribution can yield evidence of a frailty or a macro effect. Thus, we could potentially find that the GAM and GB model fit better on an aggregate level. Moreover, the function chosen here could potentially be well approximated by a step function such as the result from a GB model.

Chapter 3

Modeling Frailty Correlated Defaults with Multivariate Latent Factors

Benjamin Christoffersen and Rastin Matin

Abstract
Firm-level default models are important for bottom-up modeling of the default risk of corporate debt portfolios. However, models in the literature typically have several strict assumptions which may yield biased results, notably a linear effect of covariates on the log-hazard scale, no interactions, and the assumption of a single additive latent factor on the log-hazard scale. Using a sample of U.S. corporate firms, we provide evidence that these assumptions are too strict and matter in practice and, most importantly, we provide evidence of a time-varying effect of the relative firm size. We propose a frailty model to account for such effects that can provide forecasts for arbitrary portfolios as well. Our proposed model displays superior out-of-sample ranking of firms by their default risk and forecasts of the industry-wide default rate during the recent global financial crisis.

Keywords: frailty models, corporate default models
JEL classification: G33, G17

The authors are grateful to Mads Stenbo Nielsen, Søren Feodor Nielsen, and seminar participants at Copenhagen Business School for helpful comments.

Modeling the default risk of a corporate debt portfolio can be accomplished by modeling the default risk of the portfolio's individual firms and then aggregate up to the portfolio level. This method is advantageous as it is easy to account for changes in the portfolio through time. It is, however, commonly known that misspecification of the firm-level model or omitted variables can lead to a large downward bias in risk measures. With this in mind, we perform an analysis of a sample of large U.S. corporate firms from 1980 to 2016 and our results are twofold: First, our results strongly suggest that earlier models in the literature have been misspecified and, secondly, we present a model that accounts for the misspecification. With this model, we show that it is necessary to consider nonlinear transformations of certain variables, interactions, and account for time-varying effects of the relative firm size. Our out-of-sample results show better ranking of firms by their default risk and a better forecast of the industry-wide default rate during the last crisis.

Typical default models in the literature use firm-specific variables and macro-variables to model the probability of a future default of a firm, see e.g. Campbell et al. (2008), Chava and Jarrow (2004), Duffie et al. (2007), Shumway (2001). These models provide quite accurate ranking of firms by their default risk. However, the predicted default rate distributions of corporate debt portfolios are too narrow for some data sets and model specifications, indicating a violation of the models' assumption of conditional independence between firms given the observable variables. Many ideas have been suggested to relax this assumption (Chen and Wu, 2014, Duan and Fulop, 2013, Duffie et al., 2009, Koopman et al., 2011, 2012, Qi et al., 2014, Schwaab et al., 2017). Common for these is that they all introduce one latent variable which affects all firms equally either within an industry, rating group, or across all industries and rating groups on the log-hazard scale or logit scale in discrete time. These models are known as mixed models or frailty models when a multiplicative random factor is included in the instantaneous hazard. The addition of the random factor results in wider and more reliable prediction intervals for the default rate of a group of firms, resulting in more accurate risk measures. These models are therefore better suited for modeling risk measures of a corporate debt portfolio. Since the random factor affects groups of firms equally on the log-hazard scale, the models often do not provide better forecasts of the mean, nor do they improve the ranking of firms by their riskiness. This has been explicitly shown by Qi et al. (2014).

Within the frailty literature it is common to assume that the coefficients for observable variables are constant through time, but Lando et al. (2013) show that this assumption is too strict. Using non-parametric and semi-parametric models, they present evidence of non-constant coefficients, and not accounting for such effects may yield biased results and an invalid implied distribution for the default rate of a debt portfolio. However, the models in Lando et al. (2013) cannot directly be used for forecasting due to their non-parametric components. In this work we bridge these two approaches by presenting a frailty model that relaxes the assumption of constant coefficients, which is also able to forecast future default rates and properly account for conditional correlation.

We first ensure that any time-varying effects are not due to an invalid specification of the linear predictor. We find that additional variables are needed in the model specification of Duffie et al. (2009, 2007), since we observe a smaller difference in log-likelihood than Duffie et al. (2009) between a model with and without frailty. This is consistent with Lando and Nielsen (2010) that cannot reject the misspecification test of Duffie et al. (2007) after inclusion of additional firm-specific covariates.

Based on work by Berg (2007) and Christoffersen et al. (2019), we expect nonlinear effects of some covariates on the log-hazard scale. We account for these by natural cubic splines and, unlike Lando and Nielsen (2010), we indeed find a significant nonlinear effect of the idiosyncratic stock volatility of the firm, the net income over total assets, and log market value over total liabilities. Accounting for nonlinear effects in this manner is rarely done in the literature even though there is no a priori reason to suspect that covariates should have a linear effect on the log-hazard scale. Our findings of a nonlinear effect of the idiosyncratic stock volatility and log market value over total liabilities provide further evidence that the Merton model provides useful guidance for building default models but may need adjustments. As Filipe et al. (2016), Jensen et al. (2017), Lando et al. (2013), we also find strong evidence of a time-varying coefficient for a size measure of the firm. In this regard we note that our size variable differs from the aforementioned papers by using the market value as in Shumway (2001).

Section 3.1 describes in detail the hazard and frailty models we use, and in Section 3.2 we describe our data set. We present results for monthly hazard rate models with and without frailty in Section 3.3. We begin the section by providing evidence of nonlinear effects and an interaction in models without frailty and then we extend these models to include frailty variables. Section 3.4 contains an out-of-sample test of the models, and we conclude in Section 3.5.

3.1 Model Specification

Our baseline model is a Cox proportional hazards model with a constant baseline hazard and time-varying covariates as in Duffie et al. (2007). Thus, the conditional instantaneous hazard rate of firm i at time t is

$$\lambda_i\left(t \mid \boldsymbol{x}_i(t), \boldsymbol{m}(t)\right) = R_i(t) \exp\left(\alpha + \boldsymbol{\beta}^\top \boldsymbol{x}_i(t) + \boldsymbol{\gamma}^\top \boldsymbol{m}(t)\right) \tag{3.1}$$

where $R_i(t)$ is a censoring indicator which is zero if the individual i is censored at time t, $\boldsymbol{x}_i(t)$ are the firm-specific covariates at time t, $\boldsymbol{m}(t)$ are the macro-variables at time t and α, $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ are unknown parameters which we need to estimate. We observe $R_i(t)$, $\boldsymbol{x}_i(t)$, and $\boldsymbol{m}(t)$ at discrete points so we model these as variables that are piecewise constant. Thus, the instantaneous hazard in Equation (3.1) is piecewise constant, resulting in a piecewise exponentially distributed arrival time. Due to the discrete hazard, we simplify the notation to

$$\lambda_{ik}\left(\boldsymbol{x}_{ik}, \boldsymbol{m}_k\right) \equiv \lambda_i\left(t \mid \boldsymbol{x}_i(t), \boldsymbol{m}(t)\right) = R_{ik} \exp\left(\alpha + \boldsymbol{\beta}^\top \boldsymbol{x}_{ik} + \boldsymbol{\gamma}^\top \boldsymbol{m}_k\right) \tag{3.2}$$

where $k = \lceil t \rceil$, $\lceil t \rceil$ is the ceiling of t, and we let $\boldsymbol{x}_{ik}$ be the constant value that $\boldsymbol{x}_i(t)$ takes on the interval $(k-1, k]$ and similarly for R_{ik} and $\boldsymbol{m}_k$. Letting T_i denote the default time of firm i, then the parameters are easily estimated by using that the likelihood in discrete time (i.e., the probability of default) for firm i in period $(k-1,k]$ conditional on survival up to time $k-1$ is

$$\mathrm{P}\left(T_i \in (k-1,k] \mid T_i > k-1\right) = 1 - \exp\left(-\lambda_{ik}\left(\boldsymbol{x}_{ik}, \boldsymbol{m}_k\right)\right) \tag{3.3}$$

for $k \in \{0,1,2,\dots\}$, and in continuous time by using that the conditional density function of T_i is

$$f_i\left(k-1+\Delta t \mid T_i > k-1\right) = \lambda_{ik}\left(\boldsymbol{x}_{ik}, \boldsymbol{m}_k\right)\exp\left(-\lambda_{ik}\left(\boldsymbol{x}_{ik}, \boldsymbol{m}_k\right)\Delta t\right)$$

where $\Delta t \in (0,1)$.

The strict assumption in Equation (3.1) is that firms' default intensities, λ_{ik}, are only correlated through co-movements in firm-specific covariates, $\boldsymbol{x}_{ik}$, or the macro-variables, $\boldsymbol{m}_k$. This assumption may not hold in practice for multiple reasons: Our model could be misspecified by omission of one or more macro-variables, co-movements in an omitted firm covariate, or co-movements in a variable which is modeled incorrectly. As in Duffie et al. (2009),[1] one way to account for the excess clustering of defaults is by extending the hazard in Equation (3.2) to

$$\lambda_{ik}\left(\boldsymbol{x}_{ik}, \boldsymbol{m}_k, A_k\right) = R_{ik}\exp\left(\alpha + \boldsymbol{\beta}^\top\boldsymbol{x}_{ik} + \boldsymbol{\gamma}^\top\boldsymbol{m}_k + A_k\right) \tag{3.4}$$

$$A_k = \theta A_{k-1} + \epsilon_k, \qquad \epsilon_k \sim \mathcal{N}\left(0,\sigma^2\right) \tag{3.5}$$

where $\mathcal{N}\left(0,\sigma^2\right)$ is a normal distribution with mean zero and variance σ^2 and the innovation terms, ϵ_k, are iid. The term A_k increases the hazard of all firms in period $(k-1,k]$ by a factor $\exp(A_k)$, and such a multiplicative factor on the hazard is formally referred to as a frailty. Large values of σ increase the probability of observing larger differences in the log-hazard between consecutive periods all else being equal, while θ controls the rate of decay of the cumulated shocks. That is, the effect of the shock ϵ_k decays towards zero with the rate $\theta^{k'-k}$ for $k' > k$. The limit $\theta \to 0$ corresponds to independent values of A_k, which has previously been employed in, e.g., Christoffersen et al. (2019) as they do not find evidence of autoregressive random effects. This could possibly be due to wider time intervals and/or fewer cross sections.

The model specification determines whether A_k is a zero-mean stationary or non-stationary process. The former would be the case when the frailty captures a stationary macro-variable. However, if the model specification, e.g., includes the log of a nominal (non-real) value variable, but the true association is linear in the log of the real value, then a non-stationary adjustment to the intercept would be needed. Examples of the latter are found in Lando and Nielsen (2010) and Chava et al. (2011), where it is unclear whether the authors use nominal or real values. As we only include real values and financial ratios where inflation adjustments do not matter, we do not expect a non-stationary intercept for the aforementioned reasons.

While the model in Equation (3.4) can account for conditional correlation, it may still be a poor approximation of the true dynamics. First, although we may expect a monotone effect, it is not obvious that the log of the default intensity should have a linear dependence on the covariates. We will later account for such nonlinear effects by using natural cubic splines. Secondly, the assumption that there is only one frailty variable which affects all firms equally on the hazard scale may not be justified. A generalization is to relax the assumption of constant proportional effect of the covariates and let some of the coefficients vary over time. The resulting model is

---

[1] Duffie et al. (2009) remark that they use an Ornstein-Uhlenbeck process, but in practice they use a discrete approximation like in Equations (3.4) and (3.5).

$$\begin{aligned}\lambda_{ik}\left(\boldsymbol{x}_{ik},\boldsymbol{m}_k,A_k,\boldsymbol{B}_k,\boldsymbol{z}_{ik}\right) &= R_{ik}\exp\left(\alpha+\boldsymbol{\beta}^\top\boldsymbol{x}_{ik}+\boldsymbol{\gamma}^\top\boldsymbol{m}_k+A_k+\boldsymbol{B}_k^\top\boldsymbol{z}_{ik}\right) \qquad (3.6)\\ \begin{pmatrix}A_k\\ \boldsymbol{B}_k\end{pmatrix} &= \mathbf{F}\begin{pmatrix}A_{k-1}\\ \boldsymbol{B}_{k-1}\end{pmatrix}+\boldsymbol{\epsilon}_k, \qquad \boldsymbol{\epsilon}_k\sim\boldsymbol{\mathcal{N}}\left(\mathbf{0},\mathbf{Q}\right)\end{aligned}$$

where $\boldsymbol{z}_{ik}$ may contain a subset of the elements of $\boldsymbol{x}_{ik}$ and $\boldsymbol{\mathcal{N}}\left(\mathbf{0},\mathbf{Q}\right)$ is a multivariate normal distribution with mean vector $\mathbf{0}$ and covariance matrix $\mathbf{Q}$. The term $\boldsymbol{B}_k$ contains the random components of the slopes, and the interpretation of $\boldsymbol{B}_k$ is the change in log-hazard relative to the reference point ($\boldsymbol{B}_k=\mathbf{0}$) due to a unit increase in the covariates with a random slope. E.g., suppose that we only have one covariate in the model and it has a random slope. Then two firms i and j which differ by $x_{jk}=x_{ik}+1$ will have a relative hazard in time period $(k-1,k]$ of

$$\frac{\lambda_{jk}\left(x_{jk},\boldsymbol{m}_k,A_k,B_k,x_{jk}\right)}{\lambda_{ik}\left(x_{ik},\boldsymbol{m}_k,A_k,B_k,x_{ik}\right)}=\frac{\exp\left(\left(\beta+B_k\right)\left(x_{ik}+1\right)\right)}{\exp\left(\left(\beta+B_k\right)x_{ik}\right)}=\exp\left(\beta+B_k\right)$$

That is, B_k changes the effect of a unit change in the covariate by a factor $\exp(B_k)$.

There are many reasons to expect non-constant slopes. The accounting standards may change, banks may temporally change the way that variables affect their lending behavior, certain types of firms may be more risky in poor economic downturns, etc. One may again argue whether the frailty is a stationary process or not. If one expects that the frailty captures temporary excess default clustering, then a stationary process seems like a natural choice.

We will estimate the frailty models using a Monte Carlo expectation maximization algorithm, where the expectation step is approximated using a particle smoother. The details of the estimation are in Appendix 3.A. The software we have developed to estimate the frailty models is available at the Comprehensive R Archive Network (CRAN).

3.2 Data and Choice of Covariates

Moody's Default Risk Service Database (MDRD) is used to get the firms' default events. We define a default event as a firm which enters into either bankruptcy, bankruptcy section 77, chapter 10, chapter 11, chapter 7, or a prepackaged chapter 11. We also regard the following as default events: A distressed exchange, a dividend omission, a grace-period default, a modification of indenture, a missed interest payment and/or a missed principal payment, payment moratorium, and a suspension of payments. These events are also included by Duffie et al. (2009) from MDRD and are nearly the same events included by Lando et al. (2013). The by far most frequent event is a missed interest payment followed by a chapter 11 bankruptcy and, as some of our events are not terminal, recurrent events can occur. Firms with multiple events typically have an intermediary period in which most would consider the firm as being in a non-normal state and thus not being at risk of entering into default. Thus, we censor a firm until the resolution date provided by

MDRD or 12 months after the event if the resolution date is missing. We extend the censoring period if consecutive events fall within this default event time and the resolution date.

We only use MDRD for two reasons: First, some of the default events are closer to the point in time at which, e.g., bond holders suffer losses. Secondly, we can use the same default events for all firms in our sample. We could augment our data set with firms that are not in MDRD, but then we would track different events depending on whether the firm is tracked by Moody's. Thus, the event definition would be broader for firms in MDRD as we would likely only have legal bankruptcy events available for firms outside MDRD. Consequently, our results could reflect differences between the two groups and it would be unclear what we model.

We use CRSP and Compustat for market data and financial statements, respectively. We lag data from Compustat by 3 months to reflect the typical delay on financial statements, use quarterly data with annualized flow variables when available and otherwise we use yearly data. Data from CRSP is lagged by 1 month to reflect that we only know past market data. Summary statistics are shown in Table 3.4 in the appendix, and the firm-specific variables we include are:

- *Operating income to total assets*: Operating income after depreciation relative to total assets. It is a profitability measure and we expect that more profitable firms should be less likely to enter into default.

- *Net income to total assets*: Net income relative to total assets. It is similarly a profitability measure but includes all costs. Including both ratios allows one to distinguish between the partial association of the two types of costs.

- *Market value to total liabilities*: Market value from CRSP relative to total liabilities. A larger ratio should imply that the firm is further from default all else equal.

- *Total liabilities to total assets*: Total liabilities relative to total assets. It is an indicator of the firm's financial leverage and we expect that all else equal a higher ratio should imply a higher probability of default.

- *Current ratio*: Current assets relative to current liabilities. A too low ratio would imply that the firm may not be able meet its short-term debt obligations thus increasing the probability of default.

- *Working capital to total assets*: Working capital relative to total assets. Similar to the current ratio, it measures the ability to meet the short-term debt obligations but does so with a metric relative to the size of the firm.

- *Log current assets*: Log of current assets deflated with the U.S. Government Consumer Price Index from CRSP. This is similar to the pledgeable assets used in Lando et al. (2013) but we do not add the book value of net property, plant, and equipment to the current assets. The variable captures both the size of the firm and the assets which can be quickly converted into cash.

- *Log excess return*: 1-year lagged average of monthly log return minus the value-weighted log total market return. We require at least three months of returns. While we do not have a particular effect in mind, this variable has shown to be a strong predictor in the literature (Duffie et al., 2009, 2007, Shumway, 2001).

- *Relative log market size*: Log market value of the firm minus the log total market value. The total market value is the sum of market values of AMEX-, NYSE-, and NASDAQ-listed firms. As remarked by Shumway (2001), subtracting the log total market value from the log market value of the firm has the advantage that it deflates the nominal log market value. Low-valued firms should be closer to entering into default in which case any investments by investors are likely lost. Though, the variable also measures the size of the firm.

- *Distance to default*: Estimated 1-year distance to default. The drift and volatility of the underlying assets are estimated over the past year using the so-called KMV method as in Vassalou and Xing (2004), and we set the debt due in one year to be the short-term debt plus 50% of the long-term debt as is common. We require at least 60 days of market values to estimate the parameters.

  The statistics of our distance to default is comparable to that reported by Vassalou and Xing (2004), which is anticipated as we use the same method and listed U.S. firms.[2] However, we note that a wide range of values have been reported in the literature.[3]

- *Idiosyncratic volatility*: Estimated standard deviation of 1-year past rolling window regressions of daily log return on the value-weighted log market return. We require at least 60 days of returns in the regressions. The variable is used in Shumway (2001) and one motivation is that more volatile firms should have a higher chance of entering into default (e.g., due to more volatile cash flows as argued by Shumway, 2001).

The value-weighted market return we use is the NYSE and AMEX index from CRSP. In terms of macro-variables, we include a market return and treasury bill rate like Duffie et al. (2009, 2007), specifically the value-weighted past 1-year log return of the aforementioned index and 1-year treasury bill rate. All variables are winsorized at 1% and 99% quantile and we carry forward missing covariates for up to 3 months for CRSP-based variables and 1 year for Compustat-based variables.

All of these covariates have appeared in multiple papers before (e.g., Bharath and Shumway, 2008, Chava and Jarrow, 2004, Shumway, 2001). It is deliberate that we use covariates that have previously been used in the literature as the goal of this work is not to seek new covariates.

We include a firm in our sample as long as it is listed, we have data from Compustat and CRSP for all variables, and the firm has started being rated by Moody's or if it is less than 36 months after the rating has been withdrawn and the firm is not rated again by Moody's.

---

[2] The probabilities of default in Vassalou and Xing (2004) are available at `www.maria-vassalou.com/data/defaultdataset.zip`. Comparing our distance to default to theirs over the same period after truncating at a 10^{-15} and $1 - 10^{-15}$ probability of default as they do yields a mean and standard deviation of the distance to default of 4.856 and 2.739 , respectively. The corresponding figures in Vassalou and Xing (2004) are 4.391 and 2.608, respectively.

[3] Chava et al. (2011), Duan et al. (2012), Lando et al. (2013), Lando and Nielsen (2010), Qi et al. (2014) show a mean ranging from 1.867 to 16.79 and a standard deviation ranging from 2.653 to 12.83.

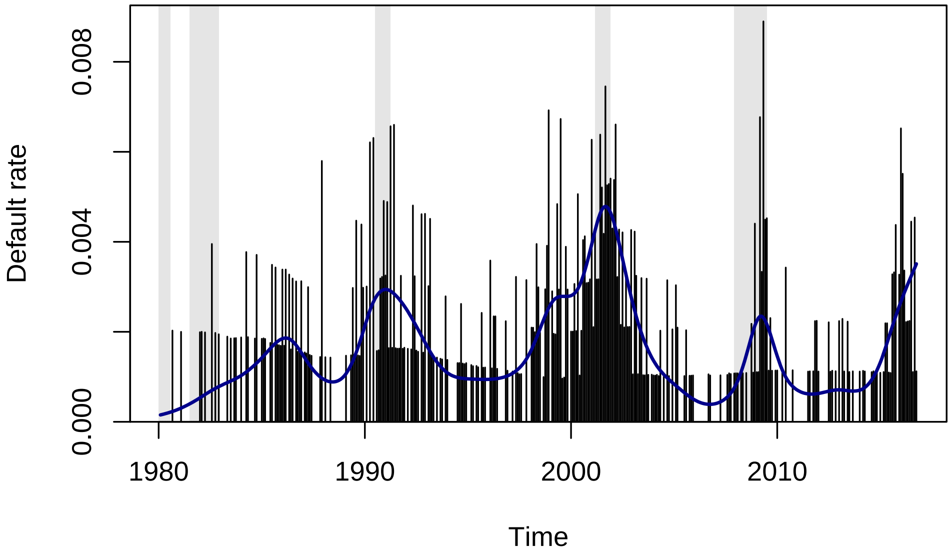


Figure 3.1: Monthly default rate in the sample. The blue line is a natural cubic spline with an integrated square second derivative cubic spline penalty. The penalty parameter is chosen using an un-biased risk estimator criterion. Gray areas are recession periods from National Bureau of Economic Research.

Firms outside this range have a virtually zero default rate in the MDRD database, which is likely because Moody's no longer tracks the firms or has not yet started to track them.

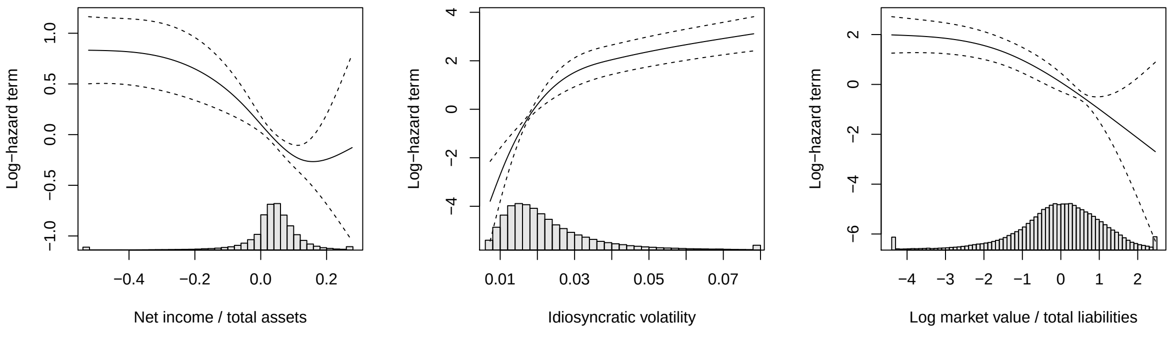
A delisting month counts as a default if a default event happens up to one year after the firm delists. This is similar to Shumway (2001) though he uses a five-year limit instead. An advantage of the event definition we use relative to, e.g., Shumway (2001) is that the events happen close to delisting or before delisting. Specifically, we observe that 84.7% of the events occur while we still have covariate information of the firm and the firm has not delisted or delists in the month of the event. Thus, we include the first half of 2016 in our out-of-sample test in Section 3.4 as we may only miss a few events since our version of MDRD was last updated in October 2016.

As is standard in the literature, we exclude firms with an SIC-code in the range 6000-6999 (financial firms) and those greater than 9000 (public administration or non-classifiable). We use the historical SIC-code from Compustat if it is available and otherwise use CRSP's historical code.

Figure 3.1 shows the monthly default rate in the sample. There is a visible clustering of defaults around economic crises, however, it is not clear from this plot whether the clustering can be captured by firm-specific variables or macro-variables.

3.3 Empirical Results

We start this section by estimating models without frailty and show that we get a better fit by adding covariates, splines, and an interaction to the model in Duffie et al. (2007). That additional covariates are needed is similar to the conclusion in Lando and Nielsen (2010), but



(a) **(b)** **(c)**

Figure 3.2: Splines in model $\mathcal{M}_3$. Splines for (a) net income to total assets, (b) the idiosyncratic volatility, and (c) log market value to total liabilities containing both the linear and nonlinear components in Table 3.1. The y-axes are the effect of the covariate on the log-hazard scale and dashed lines are 95% pointwise confidence intervals. Histograms of the covariates are shown in the background to illustrate where we have data (they have no relation to the y-axis).

treating nonlinear effects has so far received limited attention in corporate default literature. We find a nonlinear effect of variables that are related to the Merton model and discuss this finding and its relation to previous work. We then turn to frailty models, where we estimate a model with a single frailty that affects all firms equally on the log-hazard scale (cf., Equation (3.4)), and subsequently extend it to include a frailty that depends on the relative log market size of the firm as in Equation (3.6). We end by comparing our results to previous work.

Table 3.1 presents the parameter estimates for the models without frailty. Column $\mathcal{M}_1$ is similar to the specification used by Duffie et al. (2007) and Duffie et al. (2009) when the frailty variable is not included. A one standard deviation change in distance to default (log excess return) is associated with an $\exp(-2.263)$ ($\exp(-0.932)$) factor change in the hazard. The former shows that distance to default is a good predictor of default as changes in distance to default are associated with large changes in the instantaneous hazard. However, the latter shows that distance to default is not a sufficient on its own, as is also observed in Bharath and Shumway (2008).

The signs of the coefficient estimates for the two macro-variables are similar to Duffie et al. (2009) and like them we do not have an intuitive explanation for the log market return slope estimate that makes sense marginally. However, we agree that a univariate interpretation may be invalid and a plausible explanation is instead that the distance to default could be “too large” on average when entering as a linear effect on the log-hazard scale in good periods on the stock market. In this case one would expect a positive slope on the market return. We remark that our estimates are not directly comparable to those of Duffie et al. (2009, 2007) as they only (among other things) consider industrial firms, have a different time period, and include events from other databases.

The AICs of $\mathcal{M}_1$ and $\mathcal{M}_2$ show that $\mathcal{M}_2$ is a much better fit. Distance to default is not significant in model $\mathcal{M}_2$, which may not be that surprising as we include both the market value to liabilities and the idiosyncratic volatility, which have strong associations with the value and

Table 3.1: Coefficient estimates from monthly default models without frailty. All covariates are centered in the shown models. The χ^2 test statistics from the Wald tests are given in parentheses: *** implies significance at the 1% level, ** at the 5% level, and * at the 10% level. The spline rows are from a three-dimensional natural cubic spline basis, which is restricted to a two-dimensional space that is orthogonal to the linear term. Thus, the linear term can be interpreted as the linear part of the spline. The two coefficients for each basis function in the splines are omitted and a "✓" indicates that the spline term is included in the model.

	$\mathcal{M}_1$		$\mathcal{M}_2$		$\mathcal{M}_3$	
Intercept	-9.938^{***}	(2379.1)	-10.265^{***}	(1264.2)	-10.436^{***}	(911.8)
Distance to default	-0.546^{***}	(227.6)	0.001	(0.0)	0.148^{***}	(7.9)
Log excess return	-2.331^{***}	(420.7)	-1.744^{***}	(207.7)	-1.729^{***}	(205.5)
T-bill rate	1.827	(1.2)	4.459^{**}	(6.1)	4.751^{***}	(6.9)
Log market return	1.132^{***}	(18.3)	0.877^{***}	(10.0)	0.890^{***}	(10.1)
Log current assets			0.178^{***}	(18.4)	0.158^{***}	(10.7)
Working capital / total assets			-1.060^{**}	(4.1)	-0.888^{*}	(2.8)
Operating income / total assets			-1.643^{***}	(9.7)	-1.180^{**}	(4.7)
Market value / total liabilities			-1.302^{***}	(24.9)		
Log market value / total liabilities					-0.806^{***}	(11.6)
Net income / total assets			-1.156^{***}	(16.2)	-2.041^{***}	(15.2)
Total liabilities / total assets			0.595^{***}	(7.4)	0.411^{*}	(3.3)
Current ratio			-0.188^{*}	(3.7)	-0.816^{***}	(20.3)
Idiosyncratic volatility			26.844^{***}	(60.9)	108.095^{***}	(41.1)
Relative log market size			-0.356^{***}	(47.6)	-0.293^{***}	(25.0)
Net income / total assets (spline)					$✓^{**}$	(6.2)
Idiosyncratic volatility (spline)					$✓^{***}$	(27.0)
Log market value / total liabilities (spline)					$✓^{***}$	(19.3)
Current ratio · Idiosyncratic volatility					15.293^{***}	(19.8)
AIC	5511.7		5067.0		5004.4	
log-likelihood	−2750.8		−2519.5		−2481.2	
Number of firms	3020		3020		3020	

volatility of the underlying asset in the Merton model (we remark on this further in the next subsection). All signs of the coefficients are as expected except for the log current assets. As we show in the table and in Figure 3.10 in the appendix, we find that larger current assets are associated with a higher default hazard for fixed relative market size. We note that removing the relative market size from $\mathcal{M}_2$ yields a negative (but small in absolute terms) coefficient on the log current asset. Further, none of the variance inflation factors in $\mathcal{M}_2$ are larger than four, so there is no severe multicollinearity.

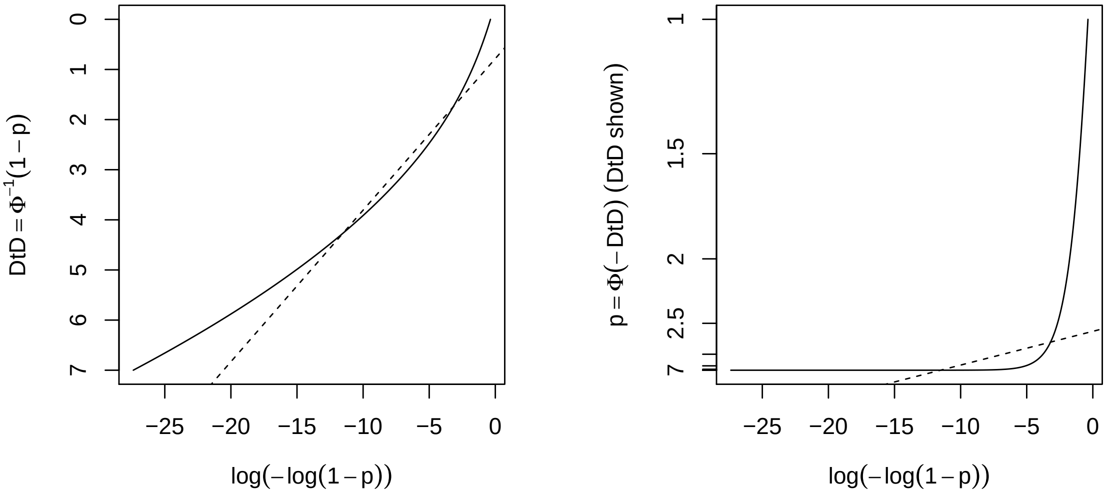
The differences between columns $\mathcal{M}_2$ and $\mathcal{M}_3$ are three spline terms and an interaction between the current ratio and the idiosyncratic volatility. The latter shows that the partial effect of the current ratio increases when the idiosyncratic volatility increases. This implies a lower partial association with a measure of capability to pay short-term debt (the current ratio) when the firm's equity value is more volatile, since the main effect is negative. This seems plausible.

The three splines in $\mathcal{M}_3$ are shown in Figure 3.2. The splines are subject to a typical sum-to-zero constraint, and a three-dimensional natural cubic spline basis is used. The left plot shows a tilted "hockey-stick" curve, which flattens for large negative loss to total assets, similar to what is observed in Christoffersen et al. (2019). The partial effect of the idiosyncratic volatility only differs for firms with small to moderate idiosyncratic volatility and flattens thereafter. Due to a relatively small number of events for firms with a moderate or large market value to total liabilities, we log-transformed the ratio to obtain a more well-posed problem. The low number of observed events in the right tail of the ratio is reflected in wider confidence bounds of the spline.

3.3.1 Distance to Default

Distance to default has received a lot of attention in the literature. Many noticeable papers cover one or more probability of default models which include distance to default in some way (Bharath and Shumway, 2008, Campbell et al., 2008, Duffie et al., 2009, 2007, Hillegeist et al., 2004). The distance to default comes from the Merton model and assumes that the underlying firm value follows a geometric Brownian motion and that the firm has issued a single zero-coupon bond. Both assumptions are potentially restrictive. In particular, the ad hoc practice of setting the debt maturing in one year to the short-term debt and 50% of the long-term debt suggests a violation of the latter assumption.

While Duffie et al. (2009, 2007) show that distance to default is a strong predictor, Campbell et al. (2008) only find smaller improvements when including distance to default in a model that also includes volatility and leverage. The latter is similar to our findings in that we cannot reject a zero slope in $\mathcal{M}_2$. Bharath and Shumway (2008) show that distance to default is not sufficient on its own and that a simpler and highly correlated metric performs equally, if not better, in ranking firms by their default risk. However, our results are not directly comparable to Bharath and Shumway (2008), as they include the probability of default from the Merton model on the log-hazard scale whereas we include the distance to default as in Duffie et al. (2009, 2007). That is, if DtD_{it} denotes the distance to default of firm i at time t and Φ is the standard normal cumulative distribution function, then Bharath and Shumway (2008) assume that the probability



(a) **(b)**

Figure 3.3: Comparison of output from the Merton model. Plots of (a) inverse standard normal cumulative distribution, Φ^{-1}, versus the complementary log-log function and (b) uniform distribution versus the complementary log-log function both for varying quantiles, p. The lines goes through the 0.00001 and 0.04 quantile to emphasize the region where we may expect to have data. The y-axis is decreasing from north to south to get a positive slope due to the negative relationship between the distance to default and the probability of default.

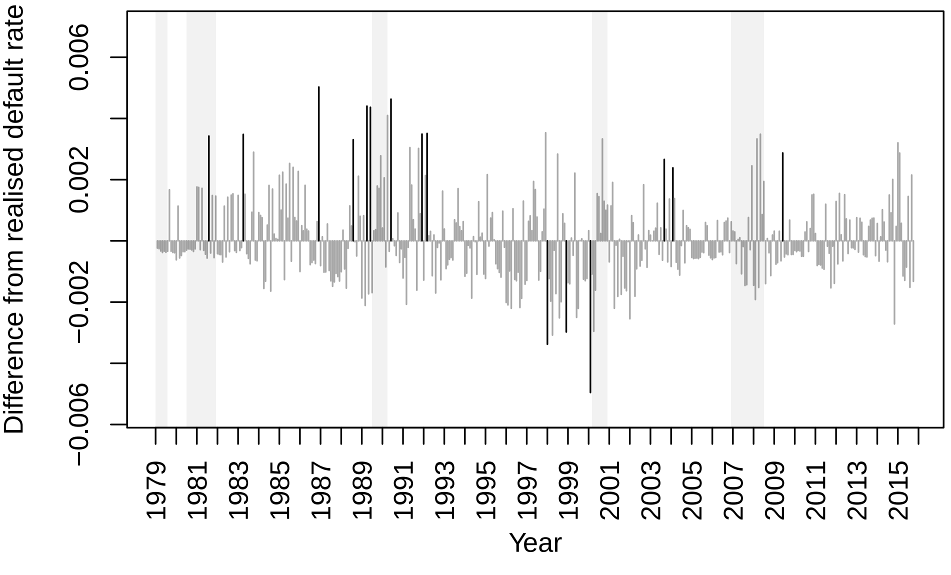


Figure 3.4: Difference between realized monthly default rate and in-sample predicted default rate of model $\mathcal{M}_3$. Black bars indicate that a 90% point-wise confidence interval does not cover the realized default rate. The scale is large compared to the monthly default rate (see Figure 3.1). Gray areas are recession periods from National Bureau of Economic Research.

of default from the Merton model

$$p_{it} = \Phi^{-1}(-\mathrm{DtD}_{it}) \tag{3.7}$$

has a linear association on the log-hazard scale while we assume that DtD_{it} has linear association on the log-hazard scale. The central question is what to expect if Equation (3.7) is approximately true.

Figure 3.3 illustrates the inverse standard normal cumulative distribution and uniform distribution versus the complementary log-log function (inverse of Equation (3.3)). The inverse standard normal cumulative distribution and complementary log-log seem to match in the low to rather high probability of default region, suggesting that including the distance to default on the log-hazard scale is not unreasonable if Equation (3.7) is approximately true. This is not the case for the uniform distribution. Moreover, we have tried to replace the distance to default in model $\mathcal{M}_1$ with the probability from Equation (3.7), resulting in an AIC of 5633 which is worse than the original model that included the distance to default. Further, performing a likelihood ratio test of the original $\mathcal{M}_1$ against the nested model which also includes the probabilities from Equation (3.7) yields a p-value of 0.083. In conclusion, we find no arguments or evidence that the probability should be used on the log-hazard scale rather than or simultaneously with distance to default.[4]

Finally, we turn to our $\mathcal{M}_3$ model. As in the online appendix of Duffie et al. (2009), we find limited evidence of a nonlinear relation on the log-hazard scale with distance to default. However, we do find a significant nonlinear relationship with two related variables, namely the idiosyncratic volatility and log market value to total liabilities. The regressions we run to estimate the idiosyncratic volatility typically do not fit well so the idiosyncratic volatility is close to the estimated volatility of the equity. The volatility of the equity is related to the volatility of underlying asset in the Merton model by

$$\begin{aligned} d_{it} &= \frac{\log V_{it}/F_{it} + (r + \sigma_{iV}^2/2)}{\sigma_{iV}} \\ \sigma_{iE} &= \frac{V_{it}}{E_{it}}\Phi(d_{it})\sigma_{iV} \end{aligned}$$

where V_{it}, E_{it}, and F_{it} is the value of the underlying asset, the value of the equity, and the value of debt maturing in one year of firm i at time t, respectively, and σ_{iE} and σ_{iV} is the volatility of the equity and underlying value of firm i, respectively. Thus, when d_{it} is large then σ_{iE} is merely a rescaling of σ_{iV}, implying that the idiosyncratic volatility is a good proxy for the underlying volatility. Consequently, adding a spline to the idiosyncratic volatility can be seen as a relaxation of the effect of one of the key components in the Merton model. Secondly, for large E_{it}/F_{it} and low σ_{iV}

$$\log\frac{V_{it}}{F_{it}} \approx \log\frac{E_{it} + F_{it}}{F_{it}} \approx \log\frac{E_{it}}{F_{it}}$$

implying that the log market value to total liabilities is a close proxy for another key component in the Merton model. Thus, two of our splines can be seen as a relaxation of the assumptions in the Merton model.

---

[4] Though, Bharath and Shumway (2008) remark that they find inferior performance by including the log of the distance to default. However, it is unclear to us how negative distance to default values are handled. It also seems to be more common to include the untransformed distance to default on the log-hazard or log odds scale.

Our choice of covariates is based on previous literature which is reflected in our preference for the idiosyncratic volatility and market value over total liabilities instead of the volatility of the equity and the market value plus total liabilities over total liabilities. All the splines we include are based on standard residual diagnostic plots. In this regard, it is interesting that we find evidence of nonlinear effects for two variables that are this closely related to the Merton model. As Bharath and Shumway (2008) conclude, the Merton model seems to provide guidance to default models but we find that relaxations seem to be needed.

To motivate the next section, Figure 3.4 shows the in-sample difference between the predicted default rate (percentage of firms that experience an event in one month) and the realized monthly rate for $\mathcal{M}_3$, and reveals that the model may have issues with excess clustering of defaults. There is a tendency of co-occurrences of too large or too small predicted monthly default rates, which suggests a frailty model with a temporal dependence as in Equation (3.4).

3.3.2 Frailty Models

We will present results for two frailty models in this section. One with a random intercept as in Equation (3.4) and one where we add a random slope for the relative log market size. Table 3.2 shows the estimated parameters. Column $\mathcal{M}_4$ shows estimates for the model with splines and the interaction ($\mathcal{M}_3$) with an added frailty effect for the intercept as in Equation (3.4). The parameter estimates are similar to $\mathcal{M}_3$. The estimated loading θ is close to what Duffie et al. (2009) find. However, the difference Δ in twice the log-likelihood is only 7.0 (the log-Bayes factor mentioned in Duffie et al., 2009). Though, Δ between the model similar to Duffie et al. (2009) ($\mathcal{M}_1$) with and without a frailty term is 14.8, which is closer to the 22.6 reported in Duffie et al. (2009). In conclusion, our finding suggests that the additional variables, nonlinear effects, and the interaction capture some of the temporal heterogeneity.

Figure 3.5 shows the predicted frailty variable, A_k, conditional on the observed data. There are some periods where the predicted value of the frailty-term is 0.2, yielding $\exp(0.2) \approx 1.22$ factor higher instantaneous hazard for all firms at the same point in time. The small difference in log-likelihood is reflected in the wide prediction intervals.

The frailty model in Equation (3.6) is denoted $\mathcal{M}_5$. We model the temporal dependence between the frailty variables as

$$\begin{pmatrix} A_k \\ B_k \end{pmatrix} = \begin{pmatrix} \theta_1 & 0 \\ 0 & \theta_2 \end{pmatrix} \begin{pmatrix} A_{k-1} \\ B_{k-1} \end{pmatrix} + \boldsymbol{\epsilon}_k, \qquad \boldsymbol{\epsilon}_k \sim \boldsymbol{\mathcal{N}}\left(\mathbf{0}, \begin{pmatrix} \sigma_1^2 & \sigma_1\sigma_2\rho \\ \sigma_1\sigma_2\rho & \sigma_2^2 \end{pmatrix}\right) \tag{3.8}$$

where B_k is the zero-mean term on the slope of the relative log market size. There is a large difference between twice the log-likelihood of model $\mathcal{M}_5$ and $\mathcal{M}_4$ of 60.0, providing strong evidence in favour of the former model.

Figure 3.6 shows the predicted value of B_k. The predicted value of A_k and B_k are very similar as θ_1 and θ_2 are almost equal and the correlation coefficient, ρ, is high. There is an increase in

Table 3.2: Estimated monthly frailty models. Both models correspond to the model with splines and an interaction, $\mathcal{M}_3$, with additional frailty variables: $\mathcal{M}_4$ has a random intercept and $\mathcal{M}_5$ has a random intercept and slope for the relative log market size. All covariates are centered in the shown models. The χ^2 test statistics from the Wald tests are given in parentheses: *** implies significance at the 1% level, ** at the 5% level, and * at the 10% level. The spline rows are from a three-dimensional natural cubic spline basis, which is restricted to a two-dimensional space that is orthogonal to the linear term. Thus, the linear term can be interpreted as the linear part of the spline. The two coefficients for each basis function in the splines are omitted and a "✓" indicates that the spline term is included in the model. The estimated splines are shown in Figure 3.11 in the appendix.

	$\mathcal{M}_4$		$\mathcal{M}_5$	
Intercept	-10.507^{***}	(878.2)	-10.789^{***}	(290.2)
Distance to default	0.137^{***}	(6.7)	0.107^{**}	(4.0)
Log excess return	-1.734^{***}	(201.8)	-1.673^{***}	(187.0)
T-bill rate	5.671^{**}	(6.3)	-0.665	(0.0)
Log market return	0.945^{***}	(7.2)	0.981^{***}	(9.4)
Log current assets	0.207^{***}	(15.1)	0.249^{***}	(23.2)
Working capital / total assets	-0.981^{*}	(3.4)	-1.096^{**}	(4.2)
Operating income / total assets	-1.073^{*}	(3.8)	-1.108^{**}	(4.1)
Log market value / total liabilities	-0.731^{***}	(9.5)	-0.633^{***}	(7.1)
Net income / total assets	-2.019^{***}	(14.9)	-1.929^{***}	(13.7)
Total liabilities / total assets	0.447^{**}	(3.8)	0.521^{**}	(5.2)
Current ratio	-0.847^{***}	(21.2)	-0.809^{***}	(19.0)
Idiosyncratic volatility	109.290^{***}	(42.1)	126.717^{***}	(48.2)
Relative log market size	-0.340^{***}	(29.1)	-0.415^{***}	(7.1)
Net income / total assets (spline)	✓*	(5.9)	✓*	(5.2)
Idiosyncratic volatility (spline)	✓***	(28.3)	✓***	(34.8)
Log market value / total liabilities (spline)	✓***	(17.3)	✓***	(18.1)
Current ratio · Idiosyncratic volatility	16.074^{***}	(21.2)	15.108^{***}	(18.6)
θ_1	0.897		0.972	
θ_2			0.974	
σ_1	0.116		0.269	
σ_2			0.064	
ρ			0.991	
AIC	5001.4		4947.4	
log-likelihood	-2477.7		-2447.7	
Number of firms	3020		3020	

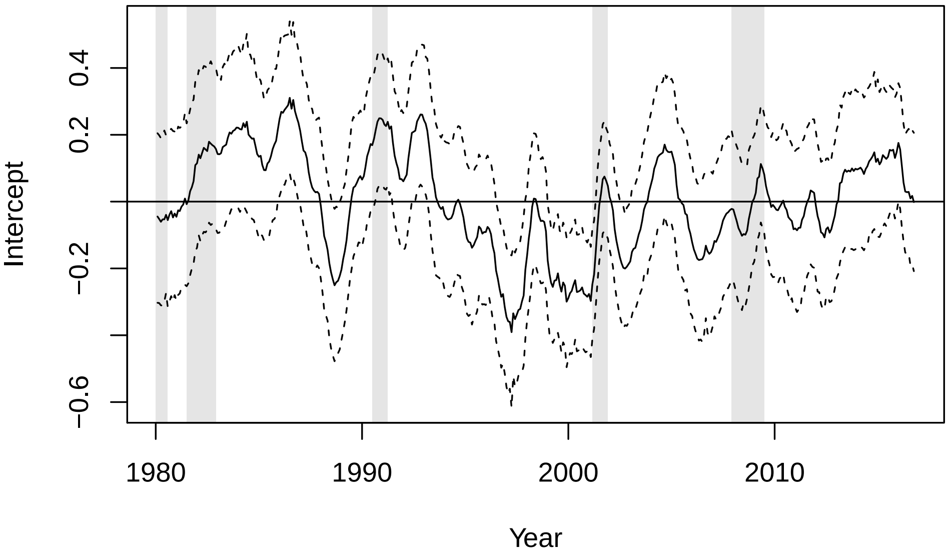


Figure 3.5: Predicted (smoothed) frailty variable A_k of $\mathcal{M}_4$ conditional on the observed data. The dashed lines are 68.3% point-wise prediction intervals (similar to the ±1 standard deviation in Duffie et al., 2009). Gray areas are recession periods from National Bureau of Economic Research.

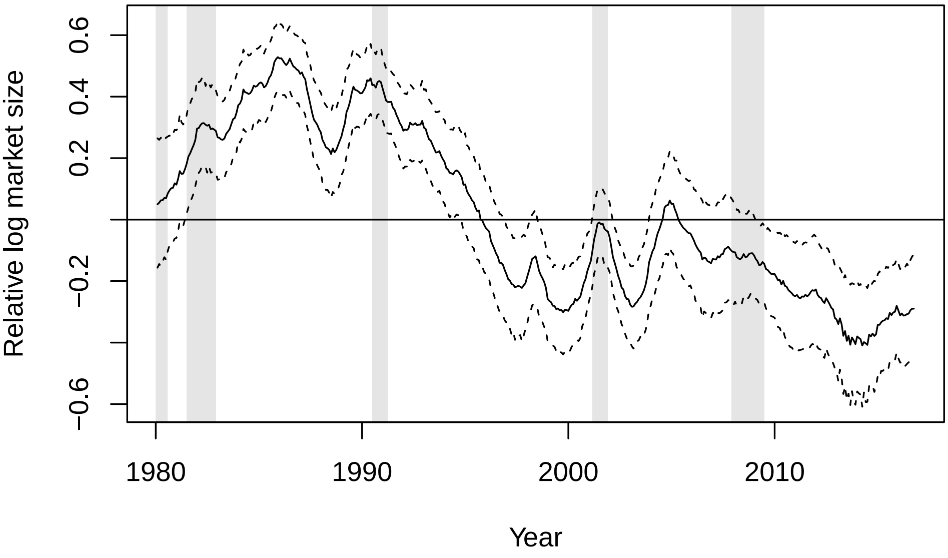


Figure 3.6: Predicted (smoothed) frailty variable B_k of $\mathcal{M}_5$ conditional on the observed data. The dashed lines are 68.3% point-wise prediction intervals (similar to the ±1 standard deviation in Duffie et al., 2009). A_k is very similar and roughly $0.264/0.064 \approx 4.1$ times greater in magnitude due to the very similar decay-rate (θ) estimate and high correlation between the two random effects. The plot does not include the fixed slope estimate, i.e., -0.415 needs to be added to get the slope on relative log market size at any point in time. Gray areas are recession periods from National Bureau of Economic Research.

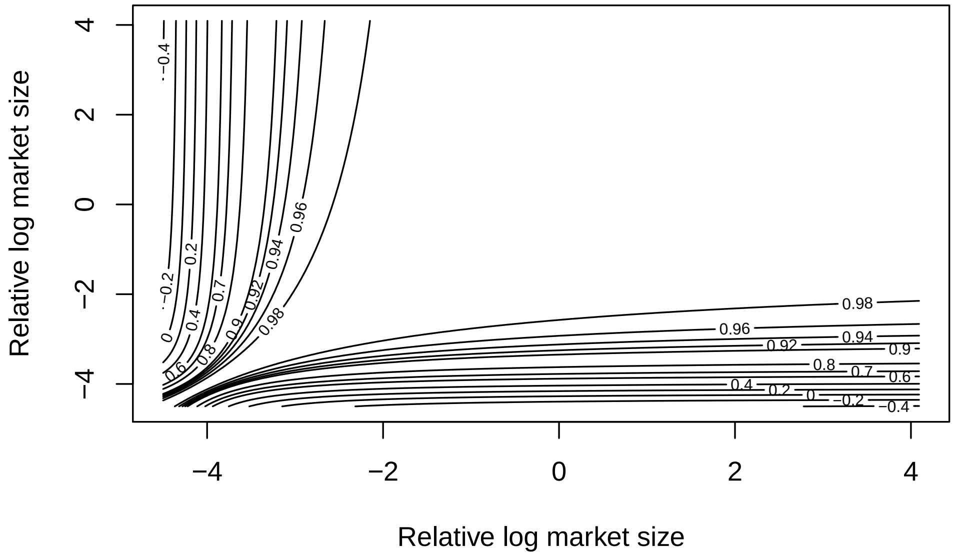


Figure 3.7: Correlation in innovation term for different values of relative log market size. Contours of correlation between $\epsilon_{1t} + \epsilon_{2t} z_{it}$ for different values of the log relative market size, z_{it}. The log relative market size is centered as in the $\mathcal{M}_5$ model.

the relative log market size slope at each of the crisis in 1990s and 2001, implying that large firms tend to be relatively more risky during a crisis all else equal. A similar time-varying effect of the size-variable is shown in Jensen et al. (2017), Lando et al. (2013) for a broad sample of Danish firms, and Filipe et al. (2016) for SMEs in Europe.[5] We remark that the size variable differs between the aforementioned papers and our work. For completeness we tried to use the log of the real value of total assets similar to Filipe et al. (2016) and Jensen et al. (2017) instead of the relative log market size in $\mathcal{M}_2$, but this resulted in a worse fit.

We note that the slope of the risk-free rate is virtually zero in model $\mathcal{M}_5$, while all other coefficients are similar to those of model $\mathcal{M}_4$. This smaller association with macro-variables is also observed in Lando et al. (2013).

Lastly, we consider the estimated random effect. Each firm's random effect component on the log-hazard scale is given by

$$A_t + B_t z_{it} = \epsilon_{1t} + \epsilon_{2t} z_{it} + \theta_1 A_{t-1} + \theta_2 B_{t-1} z_{it}$$

where z_{it} is the relative log market size. Figure 3.7 shows the correlation between the $\epsilon_{1t} + \epsilon_{2t} z_{it}$ term for two values of z_{it}. The conclusion is that firms of equal size have a highly correlated random effect term, while low-valued firms tend to have a lower correlation with the random effect term of moderately- and high-valued firms. This is in contrast to the model $\mathcal{M}_4$ where the random effect term for all firms is perfectly correlated by construction.

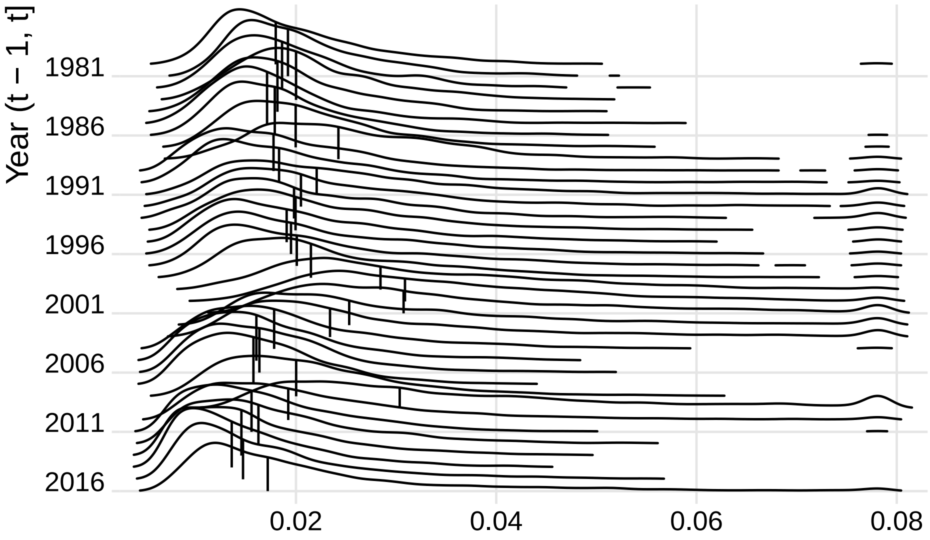


Figure 3.8: Density estimates of idiosyncratic volatility. The plot shows density estimates of the winsorized idiosyncratic volatility through time. The middle vertical lines are the medians. Densities lower than 1% of the maximum are omitted. A Gaussian kernel is used with a bandwidth of 0.00137.

3.3.3 Comparison with Other Work

Omitting covariates with co-movements through time for all firms or groups of firms can yield evidence of a single shared frailty (i.e., a time-varying intercept). An example of a covariate with a co-movement through time is shown in Figure 3.8, showing density estimates of the winsorized idiosyncratic volatility through time. Many of the covariates in our model have time-varying distributions, which may explain the smaller Bayes factor we observe between a model with a time-varying intercept versus a model without in comparison to Duffie et al. (2009). The results presented by Lando and Nielsen (2010) are yet another indication of this effect, since they fail to reject the misspecification test in Das et al. (2007) after including additional covariates.

A different approach to default modeling is to consider aggregate defaults. Some examples are Azizpour et al. (2018), Koopman et al. (2011, 2012), Schwaab et al. (2017) who aggregate to different levels, which are either total default counts or default counts in rating and industry (and region or age cohort) groups. Azizpour et al. (2018), Koopman et al. (2011), Schwaab et al. (2017) use cross-sectional averages or medians of firm-level covariates and either explicitly or implicitly assume that firms are homogeneous given these variables at the level of default that they model. This is a strict assumption and may not be valid in practice when covariates have time-varying distributions as in Figure 3.8. Time-varying distributions of firm-level covariates can yield evidence of a macro effect, frailty variable, or contagion variable as the following example will show. We have 1000 firms which we observe over 65 periods. Each firm has a randomly distributed incorporations date which is uniformly distributed on $\{0,\dots,64\}$ and a single time-

---

[5] See the log-size coefficient in the robustness check of Filipe et al. (2016).

varying covariate X_{it} which is drawn from the mixture given by

$$X_{it} \mid U_{it} = u \sim \begin{cases} \mathcal{N}\left(0,1\right) & u = 0 \\ \mathcal{N}\left(g(t),1\right) & u = 1 \end{cases}, \qquad g(t) = 4 \cdot \left| \frac{t-33}{32} \right| - 1$$

where U_{it} is Bernoulli distributed with probability 0.2. That is, the covariate is either drawn from a time-invariant distribution or from a distribution with a time-varying mean. Defaults are terminal, we observe defaults in discrete time, and define the default intensity as

$$\lambda_{it} = \exp\left(\beta_0 + \beta_1 x_{it}\right)$$

where $(\beta_0, \beta_1) = (-4, 1)$. We perform 1000 simulations and fit a model for aggregate defaults for each simulated data set using the mean at time t, $\bar{X}_t = |R_t|^{-1} \sum_{i:R_t} X_{it}$, as a covariate where R_t is the risk set at time t. Furthermore, we fit a second model where we include time as a second order polynomial. The latter model can be seen as a model which includes either macro-variables, a contagion factor, or a shared frailty variable. We reject the likelihood ratio test between the two nested models in 618 of the 1000 simulations at the conventional 5% level, which could be interpreted as evidence of a macro, a contagion, or a frailty effect when employing cross-sectional averages of firm covariates. More importantly, both models have the undesirable feature that they miss substantial cross-sectional variation, yielding incorrect results for a corporate debt portfolio with a non-random sample of the population. For completeness, Koopman et al. (2012), Schwaab et al. (2017) do state that a firm's rating may not be sufficient statistics for default. Further, there is evidence that ratings alone may be poor proxies of risk (e.g., see Hamilton and Cantor, 2004).

Our results in Figure 3.6 suggest that larger firms are partially more risky in some periods than others, whereas Azizpour et al. (2018) show that periods with large amounts of defaulted debt are followed by a higher aggregate default rate. The advantage of our model in this context is that it can be applied to an arbitrary portfolio of firms and can distinguish between an overall change in default rates and a change in default rates for a subset of firms. The latter is key, e.g., for regulators that want to evaluate the risk exposure of banks that provide loans to a subset of firms that are not a random subset of the entire population.

We remark that we do not attempt to infer causal effects. The observed frailty effect may be either "true" frailty (i.e., temporary shocks that affect all or groups of firms), a proxy for a contagion effect (i.e., the default of one firm spreads to other firms), causal associations or non-causal associations. However, we provide a model which is an accurate firm-level as well as joint default model. Such models are needed to perform bottom-up modeling of the default risk of a corporate debt portfolio. The model can easily be extended to relax the assumption that other coefficients are constant and exploit information of all defaults through time.

3.4 Out-of-sample

In this section, we will test the performance of the models out-of-sample through time. Our goal is to test how well the models perform on the firm-level and aggregate level. The former is important as we want to be able to use the models on a portfolio which contains an arbitrary subset of the firms in our sample. The latter is important as any bias at a point in time can affect the overall predicted default rate of a portfolio and subsequent modeling of other quantities using the default model.

As described in Duffie et al. (2009), our models are so-called doubly stochastic Poisson processes conditional on the frailty variables. That is, conditional on the covariates (and frailty variable path), we have piecewise exponentially distributed arrival times. However, we do not know the future covariates when we forecast apart from the value in the present month. We also do not know whether, and if so when, the firm will exit the sample due to other reasons than a default. What we will do is take both covariates and exits as given, i.e., in our risk set we include the firms up to and including the month where they exit due to a default or for other reasons. During this period we treat the firm as if it can default unconditional on that the firm will exit within our forecasting period. E.g., if a firm exits at the end of month 4 (due to exit or a default), then the firm is at risk for 4 months. This allows us to solely test our default models' performance and not how well we can model the covariates in our model.[6]

We estimate $\mathcal{M}_1$ and $\mathcal{M}_3$-$\mathcal{M}_5$ up to January or July of a given year and then use the estimated models to forecast defaults for the following half-year. We quantify the model performance in two ways: First, we use the area under the receiver operator characteristic curve (AUC). It is a commonly used metric in the default literature and the AUC allows us to assess the firm-level performance. The interpretation of AUC is the fraction of correctly ordered firms in terms of whether or not they default within the six month period. Thus, it is the probability that a random firm in our sample with a default has a higher hazard than a random firm without a default. A value of 0.5 is random guessing and a value of 1 means that all firms with a default have a higher hazard than firms without a default. We compute the AUC of each model using the mean hazard rates, which follow from the predicted default probabilities.

Secondly, we simulate the events conditional on the predicted default probabilities for each firm using the models without frailty and use these to compute the industry-wide default rate for the following half-year. Repeating this multiple times gives us a distribution for the predicted default rate, and a correctly specified model should have decent coverage of the prediction interval of this rate. To assess this we use that the upper bound of the prediction interval is similar to a 95% Value-at-Risk but for the whole industry's default rate. For the frailty models, we first have to sample a (A_k, B_k)-pair from the so-called particle cloud (see Appendix 3.A) of the last month of the estimation period, simulate $(A_{k+1}, B_{k+1}, \dots, A_{k+6}, B_{k+6})$ conditional on the sampled (A_k, B_k)-pair, and then simulate the outcomes as we do for the models without frailty conditional

---

[6] Evidence presented in Duan et al. (2012) suggests that a first order vector autoregression as in Duffie et al. (2007) may be inappropriate when modeling covariates. Modeling the high dimensional and non-fixed size set of covariate vectors (varying since the set of firms at-risk changes through time) is interesting, but not what we pursue in this paper.

Table 3.3: Number of breaches of the upper bounds of the prediction intervals in Figure 3.9(b). The first row shows the number of strictly greater realized default rates and the second row shows greater than or equal. The large difference between the two rows is due to low realized and expected default rates in some periods. The last row shows the number of half-years in our out-of-sample test.

	$\mathcal{M}_1$	$\mathcal{M}_3$	$\mathcal{M}_4$	$\mathcal{M}_5$
# breaches ($>$)	3	3	1	0
# breaches ($\geq$)	7	7	5	4
# periods	36	36	36	36

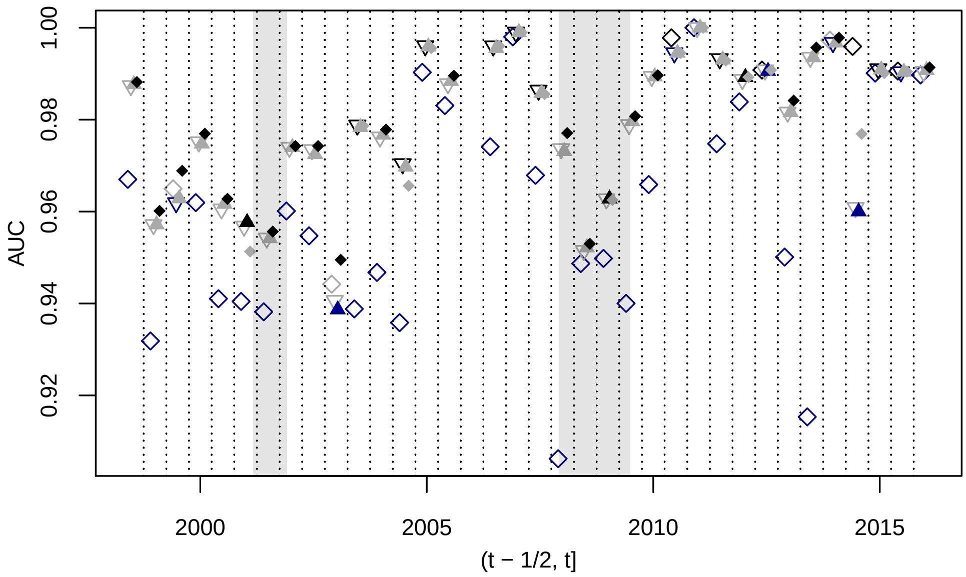
on the sampled path of the frailty variables. We do so as we need the entire paths of the frailty variables over the following six months to forecast the likelihood of an event for a given firm. We emphasize that the Monte Carlo EM algorithm, particle filter, and particle smoother only use the data available up to and including the last month of the estimation period (time k) when estimating parameters and forming the particle cloud.

The out-of-sample results are shown in Figure 3.9. The AUC in Figure 3.9(a) shows that the model like in Duffie et al. (2007) ($\mathcal{M}_1$) performs poorly in terms of firm-level performance relative to the other models. The difference in AUC sometimes exceeds 0.05 compared to the other models, which means that the latter models have more than a 5% higher fraction of correctly ordered firms by whether they default or not. This is substantial. The second conclusion is that our final frailty model ($\mathcal{M}_5$), which allows for a time-varying slope of the relative log market size, does best 19 of the 35 periods.

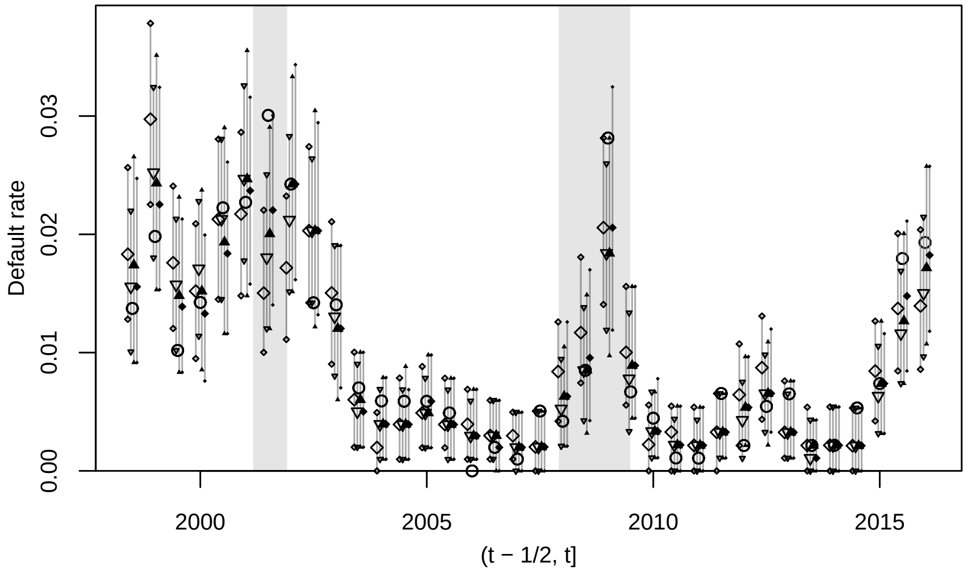
Next, we turn to the aggregate level performance. The industry-wide predicted default rates show that $\mathcal{M}_5$ is better in some periods in the sense that the median of the predicted rate is closer to the realized, and in particular we notice the crisis in 2009 and last period from January to June in 2016. However, it is not always true that the median of $\mathcal{M}_5$ is closest to the realized default rate. Table 3.3 shows the frequency of breaches of the upper bound in the prediction interval for the industry-wide default rate, and from these figures we observe that both of the two random effect models, $\mathcal{M}_4$ and $\mathcal{M}_5$, have coverage close to the 95% coverage level.

3.5 Conclusion

We have extended the hazard model of Duffie et al. (2007) by including additional covariates, a nonlinear effect for the idiosyncratic volatility, net income relative to total assets of the firm, and log market value over total liabilities, and by adding an interaction between the idiosyncratic volatility and the current ratio. This yields much better out-of-sample ranking of firms by their riskiness. Despite these additions, the model still has issues with excess default clustering although we observe less evidence for excess clustering with a random effect as in Duffie et al. (2009) compared to the model without our additions. We show that this clustering cannot be modeled by adding a single frailty effect affecting all firms equally on the log-hazard scale, as otherwise



(a)



(b)

Figure 3.9: Out-of-sample performance. Figure (a) shows the out-of-sample AUC where ⬨ is model $\mathcal{M}_1$, ▿ is $\mathcal{M}_3$, ▴ is $\mathcal{M}_4$, and ⧫ is $\mathcal{M}_5$. The points are the values over the past half-year. The vertical bars are used to separate different time periods, the model with the highest AUC is in black, the model with the lowest AUC is in blue, and the rest are gray. Results for the half-year $t = 2006$ are absent as the half-year contains no defaults. Plot (b) shows the out-of-sample default rate along with 90% prediction intervals, where the upper bound is like a 95% Value-at-Risk but for the industry default rate. The symbols in the middle denote the median value and ○ denotes the realized default rates. Gray areas are recession periods from National Bureau of Economic Research.

argued by Duffie et al. (2009).

Instead, we add a time-varying random slope to the relative log market size of the firm, similar to the size effect in Lando et al. (2013). However, unlike the semi-parametric and non-parametric models used by Lando et al. (2013), our model is an extension of previous frailty models and thus it can be used for forecasting. We show that our frailty model fits much better in-sample and, furthermore, our out-of-sample test shows superior ranking of firms by riskiness.

We also present evidence for two nonlinear effects of variables that are closely related to the Merton model. This finding corroborates the conclusion of Bharath and Shumway (2008) that the Merton model provides useful guidance for building default models but it is not sufficient.

We remark that our list of covariates may not be complete, and this may also be true for the covariates we model with nonlinear association on the log-hazard scale, included interactions and frailty variables, and the assumed distribution of frailty variables. Despite this, our study highlights that the traditional assumption of linearity on the log-hazard scale, the assumption of no interactions, and the assumption of constant slopes within corporate default modeling are too strict in our sample. With this work we show how to easily relax these assumptions in the presented models, and the software we have developed is readily available for practitioners. See the appendix for details.

Time-varying size effects like the one we show have been observed in other data sets by Filipe et al. (2016), Jensen et al. (2017), but it is yet to be determined if this is a more general effect within corporate default models.

Table 3.4: Summary statistics for the covariates used in the monthly hazard models in Section 3.3. The right-most columns show the 1% and 99% quantiles, and means and standard deviations are computed after winsorizing. The current assets are deflated by the U.S. Consumer Price Index.

	Mean	Median	Standard deviation	1%	99%
Distance to default	5.834	5.294	4.144	-1.631	18.756
Log excess return	-0.057	-0.020	0.400	-1.516	0.964
Working capital / total assets	0.153	0.129	0.164	-0.168	0.607
Operating income / total assets	0.090	0.087	0.088	-0.220	0.362
Market value / total liabilities	1.657	1.035	1.927	0.012	11.686
Net income / total assets	0.032	0.042	0.103	-0.523	0.278
Total liabilities / total assets	0.631	0.615	0.189	0.237	1.358
Current ratio	1.836	1.602	1.042	0.421	6.432
Log current assets	5.117	5.076	1.546	1.686	9.026
Idiosyncratic volatility	0.023	0.020	0.013	0.007	0.078
Relative log market size	-8.945	-8.908	1.792	-13.445	-4.852

Appendix

3.A Estimating Frailty Models

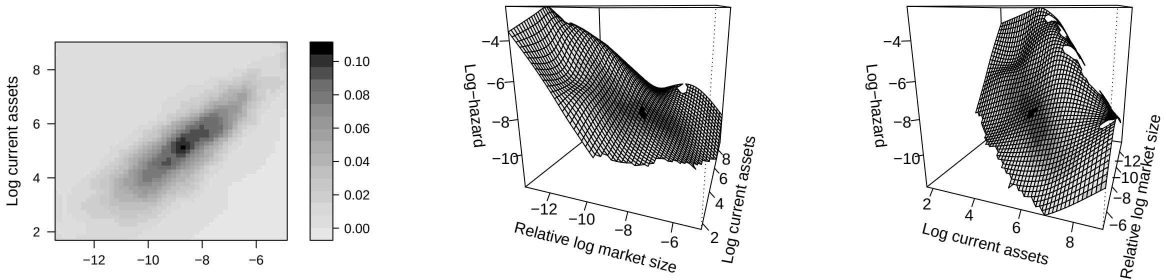
We will describe how the frailty models are estimated in this section. To do this, we start by defining the sum of the log-likelihood terms for each firm. We will focus here on the continuous case where we observe the exact event times and the most general frailty model which is shown in Equation (3.6). Let y_{ik} be one if firm i has an event in time period $(k-1,k]$ and zero otherwise. Furthermore, let Δt_{ik} be the time at risk of firm i in interval $(k-1,k]$, i.e., this will be 1 in the monthly hazard models if firm i does not have an event in the interval and otherwise the time until the event, $T_i - k + 1$. Then the log-likelihood terms from firm i conditional on the frailty variables are

$$l_i\left(A_{0:d},\boldsymbol{B}_{0:d}\right)=\sum_{k=1}^{d}\sum_{i:\,R_{ik}=1} y_{ik}\log\lambda_{ik}\left(\boldsymbol{x}_{ik},\boldsymbol{m}_k,A_k,\boldsymbol{B}_k,\boldsymbol{z}_{ik}\right)-\lambda_{ik}\left(\boldsymbol{x}_{ik},\boldsymbol{m}_k,A_k,\boldsymbol{B}_k,\boldsymbol{z}_{ik}\right)\Delta t_{ik}$$

where we observe d periods, $A_{0:d}=(A_0,A_1,\dots,A_d)$, and $\boldsymbol{B}_{0:d}=(\boldsymbol{B}_0,\boldsymbol{B}_1,\dots,\boldsymbol{B}_d)$. The complete data log-likelihood where we observe the frailty variables is

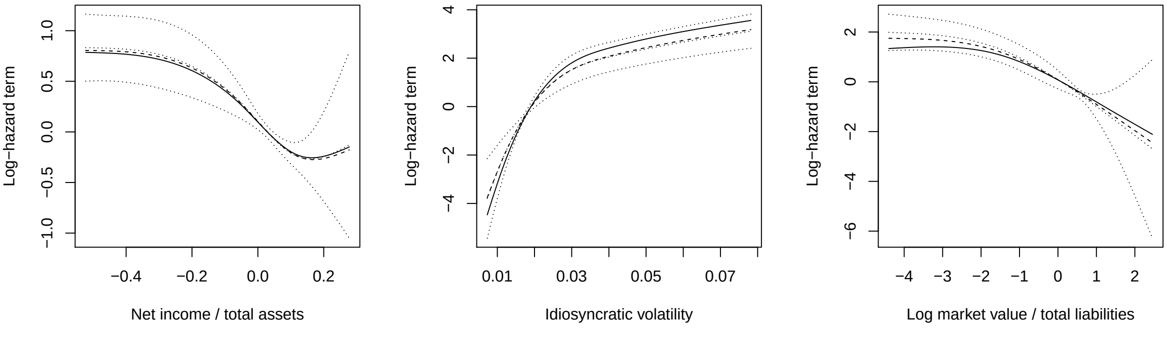
$$\mathcal{L}\left(\alpha,\boldsymbol{\beta},\boldsymbol{\gamma},\mathbf{F},\mathbf{Q}\right)=\phi\left(\begin{pmatrix}A_0\\ \boldsymbol{B}_0\end{pmatrix};\mathbf{0},\mathbf{Q}_0\right)+\sum_{k=1}^{d}\phi\left(\begin{pmatrix}A_k\\ \boldsymbol{B}_k\end{pmatrix};\mathbf{F}\begin{pmatrix}A_{k-1}\\ \boldsymbol{B}_{k-1}\end{pmatrix},\mathbf{Q}\right)+\sum_{i=1}^{n}l_i\left(A_{0:d},\boldsymbol{B}_{0:d}\right) \tag{3.9}$$

where we have n firms, $\phi\left(\cdot;\boldsymbol{v},\mathbf{V}\right)$ is the multivariate normal distribution density function with mean $\boldsymbol{v}$ and covariance matrix $\mathbf{V}$, and $\mathbf{Q}_0$ is the time-invariant covariance matrix which is given by $\mathbf{Q}_0=\mathbf{F}\mathbf{Q}_0\mathbf{F}^\top+\mathbf{Q}$. Direct maximization would require that we integrate $A_{0:d}$ and $\boldsymbol{B}_{0:d}$ out of Equation (3.9) which is infeasible. An alternative is to employ an expectation maximization (EM) algorithm (Dempster et al., 1977). This is done by starting at some value



(a) (b) (c)

Figure 3.10: Marginal effect of relative log market size and log current assets. Figure (a) shows the density of the relative log market size and log current assets. It shows that the two are correlated and the majority of the observations are along the diagonal. The density plot also shows that the two are not nearly perfectly correlated. Figures (b) and (c) show a tensor product spline from a generalized additive model with only the two variables included. The z-axis shows the log-hazard rate of default. The integral of the estimated density over the plotted surface is 99% and the colors of Figure (a) are added to the surface. We observe that along the diagonal in the xy-plane (the area where we have data) the log-hazards are higher for firms with higher current assets. The penalty parameter in the generalized additive model is chosen with an un-biased risk estimator criterion.



(a) (b) (c)

Figure 3.11: Estimated splines for the monthly frailty models. The inner dotted lines are from the model without frailty, $\mathcal{M}_3$, the dashed lines are from the model with a random intercept, $\mathcal{M}_4$, and the solid lines are from the model with a random intercept and random relative log market size slope, $\mathcal{M}_5$. The splines are for the following covariates: (a) net income to total assets, (b) the idiosyncratic volatility, and (c) log market value to total liabilities. The y-axis is the effect of the covariate on the log-hazard scale and outer dotted lines are 95% pointwise confidence intervals for the model without frailty.

$\boldsymbol{\theta}^{(0)} = \left(\alpha^{(0)}, \boldsymbol{\beta}^{(0)}, \boldsymbol{\gamma}^{(0)}, \mathbf{F}^{(0)}, \mathbf{Q}^{(0)}\right)$ and then computing

$$H\left(\boldsymbol{\theta} \,\middle|\, \boldsymbol{\theta}^{(0)}\right) = \mathrm{E}\left(\mathcal{L}\left(\alpha, \boldsymbol{\beta}, \boldsymbol{\gamma}, \mathbf{F}, \mathbf{Q}\right) \,\middle|\, \boldsymbol{y}_{1:d}, \boldsymbol{\theta}^{(0)}\right) \tag{3.10}$$

where $\boldsymbol{y}_{1:d} = (\boldsymbol{y}_1, \boldsymbol{y}_2, \dots, \boldsymbol{y}_d)$, $\boldsymbol{y}_k = (y_{1k}, y_{2k}, \dots y_{nk})$, and the expectation is w.r.t. $A_{0:d}$ and $\boldsymbol{B}_{0:d}$. This is referred to as the E-step. Then, we find a new set of parameters by

$$\boldsymbol{\theta}^{(1)} = \arg\max_{\boldsymbol{\theta}} H\left(\boldsymbol{\theta} \,\middle|\, \boldsymbol{\theta}^{(0)}\right)$$

which is referred to as the M-step. The process is then repeated with $\boldsymbol{\theta}^{(1)}$ in place of $\boldsymbol{\theta}^{(0)}$ until a convergence criterion is reached.

In our case, Equation (3.10) has no closed-form solution. Instead, we use a Monte Carlo expectation maximization algorithm where the E-step is approximated by the particle smoother suggested by Fearnhead et al. (2010). A multivariate t-distribution approximation at a mode is used at each step of the algorithm as described by Pitt and Shephard (1999). The particle smoother uses an auxiliary particle filter which is also used to get the log-likelihood approximations shown in Table 3.2. The auxiliary particle filter also yields a discrete approximation of the density of $(A_d, \boldsymbol{B}_d)$ given the observed data. The approximation is a so-called particle cloud consisting of K $(A_d, \boldsymbol{B}_d)$-pairs where each pair has a weight of its conditional probability relative to the other pairs in the cloud. The Wald tests are computed with an approximation of the observed information matrix obtained with the method suggest by Poyiadjis et al. (2011) with a particle filter as suggested by Lin et al. (2005). The R (R Core Team, 2018) package **dynamichazard** (Christoffersen, 2019) contains implementations of the methods described above and the "Particle filters in the **dynamichazard** package"-vignette in the package covers the methods in more details. The data preparation code is available at http://bit.ly/github-US_PD_data-r2 and the data analysis is available at http://bit.ly/github-US_PD_models-r3.

Chapter 4

Particle Methods in the dynamichazard Package

Benjamin Christoffersen

Abstract
This chapter introduces the particle filters and smoothers implemented in the **dynamichazard** package in R used for a class survival analysis models. In particular, the methods suggested by Briers et al. (2009), Fearnhead et al. (2010), and Poyiadjis et al. (2011) are covered in the chapter. The general methods are briefly explained towards the end whereas the majority of the chapter concerns particular application of the methods.

Keywords: survival analysis, particle filter, particle smoothing

4.1 Introduction

I will cover the implemented particle filters and smoothers in the **dynamichazard** package in this chapter along with the particle-based methods to approximate the gradient and the observed information matrix. Some of the methods and implementations are those used in the random effect model in Chapter 3. This chapter shows what is implemented, shows why, and gives a brief overview of the methods in R. Some prior knowledge of particle filters is assumed, although a brief introduction is given at the start of the chapter. Doucet and Johansen (2009) provide a tutorial on particle filters and Kantas et al. (2015) cover parameter estimation with particle filters. See also Cappé et al. (2005) for a general introduction to hidden Markov models. This chapter relies heavily on Fearnhead et al. (2010), and there is a sizable overlap between what is presented here and in the source.

The models implemented in the **dynamichazard** package are survival analysis models for terminal events. These can be in discrete time where we have binary indicators $Y_{ik} = 1_{\{T_i \in (t_{k-1}, t_k]\}}$ which is one if the random event time of individual i denoted by $T_i \in (0, \infty)$ is in the interval $(t_{k-1}, t_k]$ and zero otherwise. It can also be in continuous time where we model the distribution of the event time of individual i, T_i, with a piecewise exponential distribution conditional on observable covariates and the path of a discrete latent variable. More formally, the model is

$$\begin{aligned}
&y_{it} \sim g\left(y_{it} \mid \eta_{it}\right) & & \\
&\boldsymbol{\eta}_t = \mathbf{X}_t\boldsymbol{\alpha}_t + \boldsymbol{o}_t + \mathbf{Z}_t\boldsymbol{\omega} & & i = 1, \dots, n_t \\
&\boldsymbol{\alpha}_t = \mathbf{F}\boldsymbol{\alpha}_{t-1} + \boldsymbol{\epsilon}_t \qquad & \boldsymbol{\epsilon}_t \sim \mathcal{N}\left(\mathbf{0}, \mathbf{Q}\right), \qquad & t = 1, \dots, d \\
& & \boldsymbol{\alpha}_0 \sim \mathcal{N}\left(\boldsymbol{\mu}_0, \mathbf{Q}_0\right) &
\end{aligned} \tag{4.1}$$

where I define the conditional densities $g_t\left(\boldsymbol{y}_t \mid \boldsymbol{\alpha}_t\right) = g\left(\boldsymbol{y}_t \mid \mathbf{X}_t\boldsymbol{\alpha}_t + \boldsymbol{o}_t + \mathbf{Z}_t\boldsymbol{\omega}\right)$ and $f\left(\boldsymbol{\alpha}_t \mid \boldsymbol{\alpha}_{t-1}\right)$. For each $t = 1, \dots, d$, we have a risk set given by $R_t \subseteq \{1, 2, \dots, n\}$. Further, we let $n_t = |R_t|$ denote the number of observations at risk at time t and $n_{\max} = \max_{t \in \{1, \dots d\}} n_t$. The observed outcomes are denoted by $\boldsymbol{y}_t = \{y_{it}\}_{i \in R_t}$. $\mathbf{X}_t$ is the design matrix of the covariates and $\boldsymbol{\alpha}_t$ is the state vector containing the time-varying coefficients. The $\mathbf{Z}_t$ is the design matrix for the covariates with time-invariant coefficients and $\boldsymbol{\omega}$ represents the corresponding coefficients.

The ith row of $\mathbf{X}_t$ is $\boldsymbol{x}_{it}$, such that $\boldsymbol{x}_{it}, \boldsymbol{\epsilon}_t, \boldsymbol{\mu}_0, \boldsymbol{\alpha}_t \in \mathbb{R}^r$, $\mathbf{F}, \mathbf{Q}, \mathbf{Q}_0 \in \mathbb{R}^{r \times r}$, where the latter two are positive definite matrices, and $\boldsymbol{o}_t$s are known offsets. The data sets we are working with have $n_{\max} \gg r$ (e.g., $n_{\max} > 1000$ and $r = 5$). Thus, $n_{\max}$ is typically the primary factor for the computation time.

The **dynamichazard** package uses a particle smoother to achieve a discrete approximation of the conditional density of $\boldsymbol{\alpha}_k$ for $k = 1, \dots, d$ given the outcomes $\boldsymbol{y}_{1:d} = \{\boldsymbol{y}_1, \boldsymbol{y}_2, \dots, \boldsymbol{y}_d\}$ and uses the discrete approximation in a Monte Carlo EM-algorithm to estimate $\mathbf{Q}$, $\boldsymbol{\omega}$, and $\boldsymbol{\mu}_0$. One choice of smoother is suggested by Fearnhead et al. (2010) and another is the generalized two-filter smoother suggested by Briers et al. (2009).

The rest of the chapter is structured as follows: first, I give a brief introduction to the implemented particle filters and smoothers. Then I describe the effect of some of the arguments to particle functions in R in the package. The implemented particle filter and smoother from Fearnhead et al. (2010) is presented next, followed by the Monte Carlo EM-algorithm and the smoother suggested by Briers et al. (2009). The last section covers the implemented approximations of the gradient and observed information matrix.

4.1.1 Overview

As a brief introduction before the next sections, we will review an application of importance sampling, use this to motivate particle filtering, and then give a brief idea of the implemented particle smoothers. Suppose we want to approximate a density $c(x) = \zeta\tilde{c}(x)$, where we only know $\tilde{c}(x)$ and not the normalization constant ζ. One way to approximate the density is to

- sample $x_1, x_2, \dots, x_N$ from a distribution with density $b(x)$ where the support of b covers the support of c,
- compute the unnormalized weights $\bar{w}_i = \tilde{c}(x_i)/b(x_i)$, and
- normalize the weights $w_i = \bar{w}_i / \sum_{i=1}^N \bar{w}_i$.

This gives us the following discrete approximation of the density

$$c(x) \approx \sum_{i=1}^N w_i \delta_{x_i}(x)$$

where δ_x is the Dirac delta function which has unit point mass at x. This is directly applicable to the model in Equation (4.1) because at time 1 we want to approximate

$$p(\boldsymbol{\alpha}_1 \mid \boldsymbol{y}_1) = \frac{g_1(\boldsymbol{y}_1 \mid \boldsymbol{\alpha}_1) \int f(\boldsymbol{\alpha}_1 \mid \boldsymbol{a}_0)\, \phi(\boldsymbol{a}_0 \mid \boldsymbol{\mu}_0, \mathbf{Q}_0)\, \mathrm{d}\boldsymbol{a}_0}{p(\boldsymbol{y}_1)}$$

where $\phi(\cdot \mid \boldsymbol{m}, \mathbf{M})$ is the density function of a multivariate normal distribution with mean $\boldsymbol{m}$ and covariance matrix $\mathbf{M}$. We can easily evaluate the numerator for each $\boldsymbol{\alpha}_1$ but not the normalization constant, $p(\boldsymbol{y}_1)$.

The extension to a particle filter (which I will call a forward particle filter) is that at time 2 we want to approximate

$$p(\boldsymbol{\alpha}_{1:2} \mid \boldsymbol{y}_{1:2}) = p(\boldsymbol{\alpha}_1 \mid \boldsymbol{y}_1) \frac{g_2(\boldsymbol{y}_2 \mid \boldsymbol{\alpha}_2)\, f(\boldsymbol{\alpha}_2 \mid \boldsymbol{\alpha}_1)}{p(\boldsymbol{y}_2 \mid \boldsymbol{y}_1)}$$

Now, we can use the discrete approximation at time 1 of $p(\boldsymbol{\alpha}_1 \mid \boldsymbol{y}_1)$, sample $\boldsymbol{\alpha}_2$ given each sampled $\boldsymbol{\alpha}_1$, and apply importance sampling again. We can repeat this with similar arguments at times $3, 4, \dots, d$, reaching an approximation of $p(\boldsymbol{\alpha}_{1:d} \mid \boldsymbol{y}_{1:d})$. We will call the last element of a sampled path at time t a particle. Further, we will denote the jth particle at time t and its associated weight by $\boldsymbol{\alpha}_t^{(j)}$ and $w_t^{(j)}$ respectively.

One issue that may arise is that our samples (particles) may degenerate, so essentially only one sample path of $\boldsymbol{\alpha}_{1:d}$ has any weight in the end. To avoid this, we may introduce a resampling step. One way to resample is using the weights and letting the resampling weights be $\beta_{t+1}^{(j)} = w_t^{(j)}$, where $\beta_{t+1}^{(j)}$ is the resampling weight of particle j at time t. We then sample with replacement, using $\beta_{t+1}^{(j)}$ (or another resampling schema, so long as it satisfies certain criteria, described in e.g., Douc and Cappé, 2005). Another option when we resample the particles from time t is to use the information of the outcomes at time $t+1$, $\boldsymbol{y}_{t+1}$. This is called an auxiliary particle filter and is introduced by Pitt and Shephard (1999).

While resampling may allow for efficient use of the available hardware, it does not solve the issue of the variance of the estimate of the density of the entire path of the state vector, $\boldsymbol{\alpha}_{1:d}$. In particular, we may end up with few unique values or only one value at the early time points (say $\boldsymbol{\alpha}_1$) when we resample. Thus, it is useful to use a smoother to achieve a better approximation of the marginal density $p\left(\boldsymbol{\alpha}_t \mid \boldsymbol{y}_{1:d}\right)$. One idea in this line is to use the two-filter formula from Kitagawa (1994). However, this requires that we can evaluate $p\left(\boldsymbol{y}_{t:d} \mid \boldsymbol{\alpha}_t\right)$. It turns out that we can approximate this up to a constant, which is just what we need. This is covered in further details in Section 4.5.

The approximation uses a particle filter run backwards in time that approximates an artificial distribution. The arguments for the backward particle filter are very similar to those for the forward particle filter presented above. The kth particle in the backward particle filter at time t, its resampling weight, and the associated weight will be denoted by, respectively, $\widetilde{\boldsymbol{\alpha}}_t^{(k)}$, $\widetilde{\beta}_{t-1}^{(k)}$, and $\widetilde{w}_t^{(k)}$. The final ith smoothed particle and weight at time t will be denoted by $\widehat{\boldsymbol{\alpha}}_t^{(i)}$ and $\widehat{w}_t^{(i)}$. The latter gives us the following approximation of the marginal density of $\boldsymbol{\alpha}_t \mid \boldsymbol{y}_{1:d}$

$$p\left(\boldsymbol{\alpha}_t \mid \boldsymbol{y}_{1:d}\right) \approx \sum_{i=1}^{N_S} \widehat{w}_t^{(i)} \delta_{\widehat{\boldsymbol{\alpha}}_t^{(i)}}\left(\boldsymbol{\alpha}_t\right)$$

if we sampled N_S smoothed particles at time t. The smoothing algorithm from Fearnhead et al. (2010) is shown in Algorithm 1, the forward particle filter is shown in Algorithm 2, and the backward particle filter is shown in Algorithm 3. The arguments for the algorithms will be given later in Section 4.5. First, I will cover the choices available in the R interface of the **dynamichazard** package before covering the implemented method in the concrete application.

4.1.2 Methods in the Package

The `PF_EM` method in the **dynamichazard** package contains an implementation of the described methods. We specify the number of particles by the `N_first`, `N_fw_n_bw`, and `N_smooth` argument for N_f, N and N_s, respectively, in the Algorithm 1-3. We may want more particles in the smoothing step, $N_s > N$, as pointed out in the discussion in Fearnhead et al. (2010). Further, selecting more particles at the start of the forward and backward particle filter, $N_f > N$, may be preferable, to ensure coverage of the state space at time 0 and $d+1$.[1]

The `method` argument specifies how the filters are set up. The argument can take the following values

---

[1] We do not need to sample the time 0 and $d+1$ particles. Instead we can make a special proposal distribution for time 1 and time d. This is not implemented, though.

- `"bootstrap_filter"` for a bootstrap filter. This is where we sample using Equation (4.5), (4.12), and (4.15). This is fast, but the proposal, distribution may be a poor approximation of the distribution we want to target.

- `"PF_normal_approx_w_cloud_mean"` and `"AUX_normal_approx_w_cloud_mean"` for the Taylor approximation of the conditional density of $\boldsymbol{y}_t$ made using the mean of the parent particles and/or mean of the child particles. See Section 4.2. The `PF` and `AUX` prefixes specify whether the auxiliary version should be used.

- `"PF_normal_approx_w_particles"` and `"AUX_normal_approx_w_particles"` for the Taylor approximation of the conditional density of $\boldsymbol{y}_t$, made using the parent and/or child particle. See Section 4.2. The `PF` and `AUX` prefixes specify whether the auxiliary version should be used.

The smoother is selected with the `smoother` argument. `"Fearnhead_O_N"` gives the smoother in Algorithm 1, and `"Brier_O_N_square"` gives the smoother in Algorithm 4. The *systematic resampling* is used in all resampling steps (see Douc and Cappé, 2005, for a comparison of resampling methods). The rest of the arguments for `PF_EM` are similar to those for the `ddhazard` function covered in chapter 1 or are explained in the manual page.

It is not clear what will yield the best performance for a given data set at a fixed computation cost. One recommendation is to use the `trace` argument and check the effective sample size at each point in time during the estimation. The `"bootstrap_filter"` may not be that much cheaper in terms of computation time given that we still must evaluate g_t in Equation (4.17), (4.18), and (4.20), which have a computational complexity of $\mathcal{O}\left(n_{\max}Nr\right)$ or $\mathcal{O}\left(n_{\max}N_S r\right)$ that is typically computationally expensive given that $n_{\max}$ is large. By contrast, the `"..._w_particles"` methods have a computational complexity of $\mathcal{O}\left(n_{\max}Nr^2\right)$ or $\mathcal{O}\left(n_{\max}N_S r^2\right)$ with a potentially much larger constant. Thus, the `"..._w_cloud_mean"` may be preferred.

The rest of the chapter covers the implemented methods. It is mainly included to show exactly what is computed and why.

4.1.3 Proposal Distributions and Resampling Weights

Algorithm 1 shows one of the particle smoothers suggested by Fearnhead et al. (2010). We need to specify a series of proposal distributions and resampling weights. To show what is implemented and why, we first consider the model where

$$\boldsymbol{y}_t \mid \boldsymbol{\alpha}_t \sim \mathcal{N}\left(\mathbf{X}_t\boldsymbol{\alpha}_t + \boldsymbol{o}_t + \mathbf{Z}_t\boldsymbol{\omega}, \mathbf{H}_t\right)$$

for some known positive definite matrix $\mathbf{H}_t$. This is not implemented in this package, but deriving optimal resampling weights and proposal distributions is possible in this case. In fact, it makes little sense to use a particle filter and particle smoother because the Kalman filter and an exact smoother can be applied. However, the results here will turn out to be useful to motivate the approximations we use later. The state space model is

$$\begin{aligned}
&\boldsymbol{y}_t \sim \mathcal{N}\left(\boldsymbol{\eta}_t, \mathbf{H}_t\right) & & \\
&\boldsymbol{\eta}_t = \mathbf{X}_t\boldsymbol{\alpha}_t + \boldsymbol{o}_t + \mathbf{Z}_t\boldsymbol{\omega} & & i = 1, \dots, n_t \\
&\boldsymbol{\alpha}_t = \mathbf{F}\boldsymbol{\alpha}_{t-1} + \boldsymbol{\epsilon}_t & \boldsymbol{\epsilon}_t \sim \mathcal{N}\left(\mathbf{0}, \mathbf{Q}\right)^{,} & t = 1, \dots, d \\
& & \boldsymbol{\alpha}_0 \sim \mathcal{N}\left(\boldsymbol{\mu}_0, \mathbf{Q}_0\right) &
\end{aligned}$$

We let $\boldsymbol{h}_t = \boldsymbol{o}_t + \mathbf{Z}_t\boldsymbol{\omega}$ such that $\boldsymbol{\eta}_t = \mathbf{X}_t\boldsymbol{\alpha}_t + \boldsymbol{h}_t$ to ease the notation. We first turn to the forward particle filter in Algorithm 2. Ideally, we want the resampling weights to be

$$\begin{aligned}
\beta_t^{(j)} &\propto p\left(\boldsymbol{y}_t \,\middle|\, \boldsymbol{\alpha}_{t-1}^{(j)}\right) w_{t-1}^{(j)} \qquad (4.2)\\
&= \int g_t\left(\boldsymbol{y}_t \mid \boldsymbol{\alpha}_t\right) f\left(\boldsymbol{a}_t \,\middle|\, \boldsymbol{\alpha}_{t-1}^{(j)}\right) \mathrm{d}\boldsymbol{a}_t w_{t-1}^{(j)} \\
&= \phi\left(\boldsymbol{y}_t \,\middle|\, \mathbf{X}_t\mathbf{F}\boldsymbol{\alpha}_{t-1}^{(j)} + \boldsymbol{h}_t, \mathbf{X}_t\mathbf{Q}\mathbf{X}_t^\top + \mathbf{H}_t\right) w_{t-1}^{(j)}
\end{aligned}$$

We may notice that setting $\beta_t^{(j)} = w_{t-1}^{(j)}$ yields the so-called sequential importance resampling algorithm. For the proposal distribution, the optimal proposal density is

$$q\left(\boldsymbol{\alpha}_t \,\middle|\, \boldsymbol{\alpha}_{t-1}^{(j)}, \boldsymbol{y}_t\right) = p\left(\boldsymbol{\alpha}_t \,\middle|\, \boldsymbol{\alpha}_{t-1}^{(j)}, \boldsymbol{y}_t\right)$$

where we find that

$$\begin{aligned}
\log p\left(\boldsymbol{\alpha}_t \,\middle|\, \boldsymbol{\alpha}_{t-1}^{(j)}, \boldsymbol{y}_t\right) &= \log p\left(\boldsymbol{\alpha}_t, \boldsymbol{y}_t \,\middle|\, \boldsymbol{\alpha}_{t-1}^{(j)}\right) + \dots \\
&= \log g_t\left(\boldsymbol{y}_t \mid \boldsymbol{\alpha}_t\right) + \log f\left(\boldsymbol{\alpha}_t \,\middle|\, \boldsymbol{\alpha}_{t-1}^{(j)}\right) + \dots \\
&= -\frac{1}{2}\left(\boldsymbol{y}_t - \mathbf{X}_t\boldsymbol{\alpha}_t - \boldsymbol{h}_t\right)^\top \mathbf{H}_t^{-1}\left(\boldsymbol{y}_t - \mathbf{X}_t\boldsymbol{\alpha}_t - \boldsymbol{h}_t\right) \\
&\quad - \frac{1}{2}\left(\boldsymbol{\alpha}_t - \mathbf{F}\boldsymbol{\alpha}_{t-1}^{(j)}\right)^\top \mathbf{Q}^{-1}\left(\boldsymbol{\alpha}_t - \mathbf{F}\boldsymbol{\alpha}_{t-1}^{(j)}\right) + \dots \\
&= -\frac{1}{2}\boldsymbol{\alpha}_t^\top\boldsymbol{\Sigma}_t^{-1}\boldsymbol{\alpha}_t + \boldsymbol{\alpha}_t^\top\boldsymbol{\Sigma}_t^{-1}\boldsymbol{\mu}\left(\boldsymbol{\alpha}_{t-1}^{(j)}\right) + \dots \\
\boldsymbol{\Sigma}_t &= \left(\mathbf{Q}^{-1} + \mathbf{X}_t^\top\mathbf{H}_t^{-1}\mathbf{X}_t\right)^{-1} \qquad (4.3)\\
\boldsymbol{\mu}(\boldsymbol{x}) &= \boldsymbol{\Sigma}_t\left(\mathbf{Q}^{-1}\mathbf{F}\boldsymbol{x} + \mathbf{X}_t^\top\mathbf{H}_t^{-1}(\boldsymbol{y}_t - \boldsymbol{h}_t)\right) \qquad (4.4)
\end{aligned}$$

The ... are terms of the normalization constant. We recognize the multivariate normal distribution density, and thus the optimal proposal density is

$$q\left(\boldsymbol{\alpha}_t \,\middle|\, \boldsymbol{\alpha}_{t-1}^{(j)}, \boldsymbol{y}_t\right) = \phi\left(\boldsymbol{\alpha}_t \,\middle|\, \boldsymbol{\mu}(\boldsymbol{\alpha}_{t-1}^{(j)}), \boldsymbol{\Sigma}_t\right)$$

Alternatively, we can use the so-called bootstrap filter and let

$$q\left(\boldsymbol{\alpha}_t \,\middle|\, \boldsymbol{\alpha}_{t-1}^{(j)}, \boldsymbol{y}_t\right) = \phi\left(\boldsymbol{\alpha}_t \,\middle|\, \mathbf{F}\boldsymbol{\alpha}_{t-1}^{(j)}, \mathbf{Q}\right) \qquad (4.5)$$

which we can sample from in $\mathcal{O}\left(Nr^2\right)$ time if we have a pre-computed Cholesky decomposition of $\mathbf{Q}$. This is computationally cheap compared with the optimal solution which has a computational complexity of $\mathcal{O}\left(Nr^2 + r^3 + n_{\text{max}}r^2\right)$, but it is not optimal.

Backward filter (Algorithm 3)

We need to specify the artificial prior $\gamma_t\left(\boldsymbol{\alpha}_t\right)$ for our artificial backward distribution. Briers et al. (2009) provides recommendations for the selection. One suggestion is the artificial density function

$$\begin{aligned}
\gamma_t\left(\boldsymbol{\alpha}_t\right) &= \phi\left(\boldsymbol{\alpha}_t \,\middle|\, \overleftarrow{\boldsymbol{m}}_t, \overleftarrow{\mathbf{P}}_t\right) \\
\overleftarrow{\boldsymbol{m}}_t &= \mathbf{F}^t\boldsymbol{\mu}_0 \\
\overleftarrow{\mathbf{P}}_t &= \begin{cases} \mathbf{Q}_0 & t = 0 \\ \mathbf{F}\overleftarrow{\mathbf{P}}_{t-1}\mathbf{F}^\top + \mathbf{Q} & t > 0 \end{cases}
\end{aligned} \tag{4.6}$$

The backward arrows are added to stress that these are means and covariance matrices used the artificial distribution we target in the backward particle filter. The artificial distribution we target in backward particle filters has the following conditional density functions

$$\widetilde{p}\left(\boldsymbol{\alpha}_{t:d} \mid \boldsymbol{y}_{t:d}\right) \propto \gamma_t\left(\boldsymbol{\alpha}_t\right)\prod_{i=t}^{d} g_i\left(\boldsymbol{y}_i \mid \boldsymbol{\alpha}_i\right)\prod_{i=t}^{d-1} f\left(\boldsymbol{\alpha}_{i+1} \mid \boldsymbol{\alpha}_i\right)$$

$$\widetilde{p}\left(\boldsymbol{\alpha}_t \,\middle|\, \boldsymbol{y}_{(t+1):d}\right) \propto \gamma_t(\boldsymbol{\alpha}_t)\int \widetilde{p}\left(\boldsymbol{a}_{t+1} \,\middle|\, \boldsymbol{y}_{(t+1):d}\right)\frac{f\left(\boldsymbol{a}_{t+1} \mid \boldsymbol{\alpha}_t\right)}{\gamma_{t+1}(\boldsymbol{a}_{t+1})}\mathrm{d}\boldsymbol{a}_{t+1}$$

$$\widetilde{p}\left(\boldsymbol{\alpha}_t \mid \boldsymbol{y}_{t:d}\right) \propto g_t\left(\boldsymbol{y}_t \mid \boldsymbol{\alpha}_t\right)\widetilde{p}\left(\boldsymbol{\alpha}_t \,\middle|\, \boldsymbol{y}_{(t+1):d}\right) \tag{4.7}$$

$$\gamma_t\left(\boldsymbol{\alpha}_t \mid \boldsymbol{\alpha}_{t+1}\right) = \frac{f\left(\boldsymbol{\alpha}_{t+1} \mid \boldsymbol{\alpha}_t\right)\gamma_t(\boldsymbol{\alpha}_t)}{\gamma_{t+1}(\boldsymbol{\alpha}_{t+1})} \tag{4.8}$$

where we have left out some of the normalization constants. Sampling from this artificial distribution turns out to be useful as it gives us an approximation of a conditional density we need up to a constant (see Section 4.5). To derive the resampling weight, we first find an expression for the density $\gamma_t\left(\boldsymbol{\alpha}_t \mid \boldsymbol{\alpha}_{t+1}\right)$. We observe that

$$\begin{aligned}
\log\gamma_t\left(\boldsymbol{\alpha}_t \mid \boldsymbol{\alpha}_{t+1}\right) &= \log f\left(\boldsymbol{\alpha}_{t+1} \mid \boldsymbol{\alpha}_t\right) + \log\gamma_t(\boldsymbol{\alpha}_t) + \dots \\
&= -\frac{1}{2}\boldsymbol{\alpha}_t^\top\overleftarrow{\mathbf{S}}_t^{-1}\boldsymbol{\alpha}_t - \boldsymbol{\alpha}_t^\top\overleftarrow{\mathbf{S}}_t^{-1}\overleftarrow{\boldsymbol{a}}_t(\boldsymbol{\alpha}_{t+1}) + \dots \\
\overleftarrow{\mathbf{S}}_t &= \left(\mathbf{P}_t^{-1} + \mathbf{F}^\top\mathbf{Q}^{-1}\mathbf{F}\right)^{-1} \\
\overleftarrow{\boldsymbol{a}}_t(\boldsymbol{x}) &= \overleftarrow{\mathbf{S}}_t\left(\mathbf{P}_t^{-1}\boldsymbol{m}_t + \mathbf{F}^\top\mathbf{Q}^{-1}\boldsymbol{x}\right)
\end{aligned}$$

so

$$\gamma_t\left(\boldsymbol{\alpha}_t \mid \boldsymbol{\alpha}_{t+1}\right) = \phi\left(\boldsymbol{\alpha}_t \,\middle|\, \overleftarrow{\boldsymbol{a}}_t(\boldsymbol{\alpha}_{t+1}), \overleftarrow{\mathbf{S}}_t\right) \tag{4.9}$$

.

As in Fearnhead et al. (2010), we can show that

$$\begin{aligned}\overleftarrow{\mathbf{S}}_t &= \overleftarrow{\mathbf{P}}_t\mathbf{F}^\top\overleftarrow{\mathbf{P}}_{t+1}^{-1}\mathbf{Q}\mathbf{F}^{-\top} \\ \overleftarrow{\boldsymbol{a}}_t(\boldsymbol{x}) &= \overleftarrow{\mathbf{P}}_t\mathbf{F}^\top\overleftarrow{\mathbf{P}}_{t+1}^{-1}\boldsymbol{x} + \overleftarrow{\mathbf{S}}_t\overleftarrow{\mathbf{P}}_t^{-1}\overleftarrow{\boldsymbol{m}}_t\end{aligned} \tag{4.10}$$

e.g., by

$$\begin{aligned}&\left(\overleftarrow{\mathbf{P}}_t\mathbf{F}^\top\overleftarrow{\mathbf{P}}_{t+1}^{-1}\mathbf{Q}\mathbf{F}^{-\top}\right)^{-1}\left(\mathbf{P}_t^{-1} + \mathbf{F}^\top\mathbf{Q}^{-1}\mathbf{F}\right)^{-1} \\ &\qquad = \mathbf{F}^\top\mathbf{Q}^{-1}\overleftarrow{\mathbf{P}}_{t+1}\mathbf{F}^{-\top}\overleftarrow{\mathbf{P}}_t^{-1}\left(\overleftarrow{\mathbf{P}}_t - \overleftarrow{\mathbf{P}}_t\mathbf{F}^\top\left(\mathbf{Q} + \mathbf{F}\overleftarrow{\mathbf{P}}_t\mathbf{F}^\top\right)^{-1}\mathbf{F}\overleftarrow{\mathbf{P}}_t\right) \\ &\qquad = \mathbf{F}^\top\mathbf{Q}^{-1}\overleftarrow{\mathbf{P}}_{t+1}\mathbf{F}^{-\top} - \mathbf{F}^\top\mathbf{Q}^{-1}\mathbf{F}\overleftarrow{\mathbf{P}}_t \\ &\qquad = \mathbf{F}^\top\mathbf{Q}^{-1}\left(\mathbf{F}\overleftarrow{\mathbf{P}}_t\mathbf{F}^\top + \mathbf{Q}\right)\mathbf{F}^{-\top} - \mathbf{F}^\top\mathbf{Q}^{-1}\mathbf{F}\overleftarrow{\mathbf{P}}_t \\ &\qquad = \mathbf{I}\end{aligned}$$

where we assume that all matrices are non-singular and we use the Woodbury matrix identity. Similar arguments can be used for $\overleftarrow{\boldsymbol{a}}_t(\boldsymbol{x})$. Using the above, we find that the optimal resampling weights are

$$\begin{aligned}\widetilde{\beta}_t^{(k)} &\propto \widetilde{p}\left(\boldsymbol{y}_t \,\middle|\, \widetilde{\boldsymbol{\alpha}}_{t+1}^{(k)}\right)\widetilde{w}_{t+1}^{(k)} \\ &\propto \int g_t\left(\boldsymbol{y}_t \mid \boldsymbol{a}_t\right)\widetilde{p}\left(\boldsymbol{a}_t \,\middle|\, \widetilde{\boldsymbol{\alpha}}_{t+1}^{(k)}\right)\mathrm{d}\boldsymbol{a}_t\widetilde{w}_{t+1}^{(k)} \\ &= \phi\left(\boldsymbol{y}_t \,\middle|\, \mathbf{X}_t\overleftarrow{\boldsymbol{a}}_t(\widetilde{\boldsymbol{\alpha}}_{t+1}^{(k)}) + \boldsymbol{h}_t, \mathbf{X}_t\overleftarrow{\mathbf{S}}_t\mathbf{X}_t^\top + \mathbf{H}_t\right)\widetilde{w}_{t+1}^{(k)}\end{aligned} \tag{4.11}$$

Alternatively, we can set the resampling weights to $\widetilde{\beta}_t^{(k)} = \widetilde{w}_{t+1}^{(k)}$ and arrive at an algorithm like the sequential importance resampling algorithm. As for the proposal distribution, the optimal density is

$$\begin{aligned}\log\widetilde{p}\left(\boldsymbol{\alpha}_t \,\middle|\, \boldsymbol{y}_t, \widetilde{\boldsymbol{\alpha}}_{t+1}^{(k)}\right) &= \log\gamma_t\left(\boldsymbol{\alpha}_t\right) + \log g_t\left(\boldsymbol{y}_t \mid \boldsymbol{\alpha}_t\right) + \log f\left(\boldsymbol{\alpha}_{t+1}^{(k)} \,\middle|\, \boldsymbol{\alpha}_t\right) + \dots \\ &= -\frac{1}{2}\boldsymbol{\alpha}_t^\top\overleftarrow{\boldsymbol{\Sigma}}_t^{-1}\boldsymbol{\alpha}_t + \boldsymbol{\alpha}_t^\top\overleftarrow{\boldsymbol{\Sigma}}_t^{-1}\overleftarrow{\boldsymbol{\mu}}_t(\widetilde{\boldsymbol{\alpha}}_{t+1}^{(k)}) + \dots \\ \overleftarrow{\boldsymbol{\Sigma}}_t &= \left(\mathbf{P}_t^{-1} + \mathbf{F}^\top\mathbf{Q}^{-1}\mathbf{F} + \mathbf{X}_t^\top\mathbf{H}_t^{-1}\mathbf{X}_t\right)^{-1} \\ \overleftarrow{\boldsymbol{\mu}}_t(\boldsymbol{x}) &= \boldsymbol{\Sigma}_t\left(\mathbf{P}_t^{-1}\boldsymbol{m}_t + \mathbf{F}^\top\mathbf{Q}^{-1}\boldsymbol{x} + \mathbf{X}_t^\top\mathbf{H}_t^{-1}(\boldsymbol{y}_t - \boldsymbol{h}_t)\right)\end{aligned}$$

Thus, we set

$$\widetilde{q}\left(\boldsymbol{\alpha}_t \,\middle|\, \boldsymbol{y}_t, \widetilde{\boldsymbol{\alpha}}_{t+1}^{(k)}\right) = \phi\left(\boldsymbol{\alpha}_t \,\middle|\, \overleftarrow{\boldsymbol{\mu}}_t(\widetilde{\boldsymbol{\alpha}}_{t+1}^{(k)}), \overleftarrow{\boldsymbol{\Sigma}}_t\right)$$

A computationally simpler but nonoptimal option is to use a method like the bootstrap filter, setting

$$\widetilde{q}\left(\boldsymbol{\alpha}_t \,\middle|\, \boldsymbol{y}_t, \widetilde{\boldsymbol{\alpha}}_{t+1}^{(k)}\right) = \widetilde{p}\left(\boldsymbol{\alpha}_t \,\middle|\, \widetilde{\boldsymbol{\alpha}}_{t+1}^{(k)}\right) \tag{4.12}$$

Combining / smoothing (Algorithm 1)

We end this example with the conditional Gaussian observable outcome model with the proposal distribution needed for Algorithm 1. We want to select

$$\begin{aligned} q\left(\boldsymbol{\alpha}_t \,\middle|\, \boldsymbol{\alpha}_{t-1}^{(j)}, \boldsymbol{y}_t, \widetilde{\boldsymbol{\alpha}}_{t+1}^{(k)}\right) &= p\left(\boldsymbol{\alpha}_t \,\middle|\, \boldsymbol{\alpha}_{t-1}^{(j)}, \boldsymbol{y}_t, \widetilde{\boldsymbol{\alpha}}_{t+1}^{(k)}\right) \\ &\propto g\left(\boldsymbol{y}_t \mid \boldsymbol{\alpha}_t\right) f\left(\boldsymbol{\alpha}_t \,\middle|\, \boldsymbol{\alpha}_{t-1}^{(j)}\right) f\left(\widetilde{\boldsymbol{\alpha}}_{t+1}^{(k)} \,\middle|\, \boldsymbol{\alpha}_t\right) \end{aligned}$$

Looking at the log density as we did before, we find that

$$\begin{aligned} \log q\left(\boldsymbol{\alpha}_t \,\middle|\, \boldsymbol{\alpha}_{t-1}^{(j)}, \boldsymbol{y}_t, \widetilde{\boldsymbol{\alpha}}_{t+1}^{(k)}\right) \\ &= -\frac{1}{2}\boldsymbol{\alpha}_t^\top \overleftrightarrow{\boldsymbol{\Sigma}}_t^{-1}\boldsymbol{\alpha}_t + \boldsymbol{\alpha}_t^\top \overleftrightarrow{\boldsymbol{\Sigma}}_t^{-1}\overleftrightarrow{\boldsymbol{\mu}}_t(\boldsymbol{\alpha}_{t-1}^{(j)}, \widetilde{\boldsymbol{\alpha}}_{t+1}^{(k)}) + \dots \\ \overleftrightarrow{\boldsymbol{\Sigma}}_t &= \left(\mathbf{Q}^{-1} + \mathbf{F}^\top\mathbf{Q}^{-1}\mathbf{F} + \mathbf{X}_t^\top\mathbf{H}^{-1}\mathbf{X}_t\right)^{-1} \qquad (4.13) \\ \overleftrightarrow{\boldsymbol{\mu}}_t(\boldsymbol{x}, \overleftarrow{\boldsymbol{x}}) &= \overleftrightarrow{\boldsymbol{\Sigma}}_t\left(\mathbf{Q}^{-1}\mathbf{F}\boldsymbol{x} + \mathbf{F}^\top\mathbf{Q}^{-1}\overleftarrow{\boldsymbol{x}} + \mathbf{X}_t^\top\mathbf{H}_t^{-1}(\boldsymbol{y}_t - \boldsymbol{h}_t)\right) \qquad (4.14) \end{aligned}$$

so that

$$q\left(\boldsymbol{\alpha}_t \,\middle|\, \boldsymbol{\alpha}_{t-1}^{(j)}, \boldsymbol{y}_t, \widetilde{\boldsymbol{\alpha}}_{t+1}^{(k)}\right) = \phi\left(\boldsymbol{\alpha}_t \,\middle|\, \overleftrightarrow{\boldsymbol{\mu}}_t(\boldsymbol{\alpha}_{t-1}^{(j)}, \widetilde{\boldsymbol{\alpha}}_{t+1}^{(k)}), \boldsymbol{\Sigma}_t\right)$$

Alternatively, we can use a method like the bootstrap filter with a proposal distribution with

$$\begin{aligned} \overleftrightarrow{\boldsymbol{\Sigma}}_t &= \left(\mathbf{Q}^{-1} + \mathbf{F}^\top\mathbf{Q}^{-1}\mathbf{F}\right)^{-1} \\ \overleftrightarrow{\boldsymbol{\mu}}_t(\boldsymbol{x}, \overleftarrow{\boldsymbol{x}}) &= \boldsymbol{\Sigma}_t\left(\mathbf{Q}^{-1}\mathbf{F}\boldsymbol{x} + \mathbf{F}^\top\mathbf{Q}^{-1}\overleftarrow{\boldsymbol{x}}\right) \end{aligned} \qquad (4.15)$$

This is not optimal but faster.

Algorithm 1 $\mathcal{O}(N)$ particle smoother using the method suggested by Fearnhead et al. (2010).

```
Input:
    $\mathbf{Q}, \mathbf{Q}_0, \boldsymbol{a}_0, \mathbf{X}_1, \dots, \mathbf{X}_d, \mathbf{Z}_1, \dots, \mathbf{Z}_d, \boldsymbol{o}_1, \dots, \boldsymbol{o}_d, \boldsymbol{y}_1, \dots, \boldsymbol{y}_d, R_1, \dots, R_d, \boldsymbol{\omega}$
    Proposal distribution with density
        $q\left(\boldsymbol{\alpha}_t \;\middle|\; \boldsymbol{\alpha}_{t-1}^{(j)}, \boldsymbol{y}_t, \widetilde{\boldsymbol{\alpha}}_{t+1}^{(k)}\right) \qquad (4.16)$
procedure Filter forward
    Run a forward particle filter to yield particle clouds $\left\{\boldsymbol{\alpha}_t^{(j)}, w_t^{(j)}, \beta_{t+1}^{(j)}\right\}_{j=1,\dots,N}$ approximating the density $p\left(\boldsymbol{\alpha}_t \mid \boldsymbol{y}_{1:t}\right)$ for $t = 0, 1, \dots, d$. See Algorithm 2.
procedure Filter backwards
    Run a similar backward filter to yield $\left\{\widetilde{\boldsymbol{\alpha}}_t^{(k)}, \widetilde{w}_t^{(k)}, \widetilde{\beta}_{t-1}^{(k)}\right\}_{k=1,\dots,N}$ approximating the artificial density $\widetilde{p}\left(\boldsymbol{\alpha}_t \mid \boldsymbol{y}_{t:d}\right)$ for $t = d+1, d, d-1, \dots, 1$. See Algorithm 3.
procedure Smooth (combine)
    for $t = 1, \dots, d$ do
        Resample
        Sample $i = 1, 2, \dots, N_s$ pairs of $(j_i, k_i) \in N^2$ where each component is independently sampled using resampling weights $\beta_t^{(j)}$ and $\widetilde{\beta}_t^{(k)}$.
        Propagate
        Sample particles $\widehat{\boldsymbol{\alpha}}_t^{(i)}$ from the proposal distribution $\widetilde{q}\left(\cdot \;\middle|\; \boldsymbol{\alpha}_{t-1}^{(j_i)}, \boldsymbol{y}_t, \widetilde{\boldsymbol{\alpha}}_{t+1}^{(k_i)}\right)$.
        Re-weight
        Assign each particle a weight
            $\widehat{w}_t^{(i)} \propto \frac{f\left(\widehat{\boldsymbol{\alpha}}_t^{(i)} \;\middle|\; \boldsymbol{\alpha}_{t-1}^{(j_i)}\right) g_t\left(\boldsymbol{y}_t \;\middle|\; \widehat{\boldsymbol{\alpha}}_t^{(i)}\right) f\left(\widetilde{\boldsymbol{\alpha}}_{t+1}^{(k_i)} \;\middle|\; \widehat{\boldsymbol{\alpha}}_t^{(i)}\right) w_{t-1}^{(j_i)} \widetilde{w}_{t+1}^{(k_i)}}{\widetilde{q}\left(\widehat{\boldsymbol{\alpha}}_t^{(i)} \;\middle|\; \boldsymbol{\alpha}_{t-1}^{(j_i)}, \boldsymbol{y}_t, \widetilde{\boldsymbol{\alpha}}_{t+1}^{(k_i)}\right) \beta_t^{(j_i)} \widetilde{\beta}_t^{(k_i)} \gamma_{t+1}\left(\widetilde{\boldsymbol{\alpha}}_{t+1}^{(k_i)}\right)} \qquad (4.17)$
```

Algorithm 2 Forward filter as in Pitt and Shephard (1999). The version and notation below is from Fearnhead et al. (2010).

```
Input:
    Proposal distribution with density
        $q\left(\boldsymbol{\alpha}_t \,\middle|\, \boldsymbol{\alpha}_{t-1}^{(j)}, \boldsymbol{y}_t\right)$
    Function $h$ to compute resampling weights
        $\beta_t^{(j)} \propto h(\boldsymbol{y}_t, \boldsymbol{\alpha}_{t-1}^{(j)}) w_{t-1}^{(j)}$
Sample $\boldsymbol{\alpha}_0^{(1)}, \dots, \boldsymbol{\alpha}_0^{(N_f)}$ particles from $\mathcal{N}(\boldsymbol{\mu}_0, \mathbf{Q}_0)$ and set the weights $w_0^{(1)}, \dots, w_0^{(N_f)}$ to $1/N_f$.
for $t = 1, \dots, d$ do
    procedure Resample
        Compute resampling weights $\beta_t^{(j)}$ using $h$ and resample according to $\beta_t^{(j)}$ to yield indices $j_1, \dots j_N$. If we do not resample, then set $\beta_t^{(j)} = 1/N$ or $1/N_f$ at time $t = 1$.
    procedure Propagate
        Sample new particles $\boldsymbol{\alpha}_t^{(i)}$ using the proposal distribution $q\left(\boldsymbol{\alpha}_t \,\middle|\, \boldsymbol{\alpha}_{t-1}^{(j_i)}, \boldsymbol{y}_t\right)$.
    procedure Reweight
        Reweight particles using
            $w_t^{(i)} \propto \frac{g_t\left(\boldsymbol{y}_t \,\middle|\, \boldsymbol{\alpha}_t^{(i)}\right) f\left(\boldsymbol{\alpha}_t^{(i)} \,\middle|\, \boldsymbol{\alpha}_{t-1}^{(j_i)}\right) w_{t-1}^{(j_i)}}{q\left(\boldsymbol{\alpha}_t^{(i)} \,\middle|\, \boldsymbol{\alpha}_{t-1}^{(j_i)}, \boldsymbol{y}_t\right) \beta_t^{(j_i)}} \qquad (4.18)$
```

Algorithm 3 Backwards filter. See Briers et al. (2009) and Fearnhead et al. (2010).

```
Input:
    An artificial distribution
        $\widetilde{p}\left(\boldsymbol{\alpha}_t \mid \boldsymbol{y}_{t:d}\right) \propto \gamma_t\left(\boldsymbol{\alpha}_t\right) p\left(\boldsymbol{y}_{t:d} \mid \boldsymbol{\alpha}_t\right)$    (4.19)
    with an artificial prior distribution $\gamma_t\left(\boldsymbol{\alpha}_t\right)$.
    Proposal distribution
        $\widetilde{q}\left(\boldsymbol{\alpha}_t \,\middle|\, \boldsymbol{y}_t, \widetilde{\boldsymbol{\alpha}}_{t+1}^{(k)}\right)$
    Function $h$ to compute resampling weights
        $\widetilde{\beta}_t^{(k)} \propto h(\boldsymbol{y}_t, \widetilde{\boldsymbol{\alpha}}_{t+1}^{(k)}) \widetilde{w}_{t+1}^{(k)}$
Sample $\widetilde{\boldsymbol{\alpha}}_{d+1}^{(1)}, \dots, \widetilde{\boldsymbol{\alpha}}_{d+1}^{(N_f)}$ particles from $\gamma_{d+1}(\cdot)$ and set the weights $\widetilde{w}_{d+1}^{(1)}, \dots, w_{d+1}^{(N_f)}$ to $1/N_f$.
for $t = d, \dots, 1$ do
    procedure RESAMPLE
        Compute resampling weights $\widetilde{\beta}_t^{(k)}$ using $h$ and resample according to $\widetilde{\beta}_t^{(k)}$ to yield indices $k_1, \dots k_N$. If we do not resample, then set $\widetilde{\beta}_t^{(k)} = 1/N$ or $1/N_f$ at time $t = d$.
    procedure PROPAGATE
        Sample new particles $\widetilde{\boldsymbol{\alpha}}_t^{(i)}$ using the proposal distribution $\widetilde{q}\left(\boldsymbol{\alpha}_t \,\middle|\, \widetilde{\boldsymbol{\alpha}}_{t+1}^{(k_i)}, \boldsymbol{y}_t\right)$.
    procedure REWEIGHT
        Reweight particles using
            $\widetilde{w}_t^{(i)} \propto \frac{g_t\left(\boldsymbol{y}_t \,\middle|\, \widetilde{\boldsymbol{\alpha}}_t^{(i)}\right) f\left(\widetilde{\boldsymbol{\alpha}}_{t+1}^{(k_i)} \,\middle|\, \widetilde{\boldsymbol{\alpha}}_t^{(i)}\right) \gamma_t\left(\widetilde{\boldsymbol{\alpha}}_t^{(i)}\right) \widetilde{w}_{t+1}^{(k_i)}}{q\left(\widetilde{\boldsymbol{\alpha}}_t^{(i)} \,\middle|\, \widetilde{\boldsymbol{\alpha}}_{t+1}^{(k_i)}, \boldsymbol{y}_t\right) \gamma_{t+1}\left(\widetilde{\boldsymbol{\alpha}}_{t+1}^{(k_i)}\right) \beta_t^{(k_i)}}$    (4.20)
```

4.2 Nonlinear Conditional Observation Model

If we go back to the model in Equation (4.1), then $\boldsymbol{y}_t \mid \boldsymbol{\alpha}_t$ is not a multivariate normal distribution for the implemented models, and we have no closed-form solutions for the optimal resampling weights. We also do not know the following conditional distributions: $\boldsymbol{\alpha}_t \mid \boldsymbol{y}_t, \boldsymbol{\alpha}_{t-1}$, $\boldsymbol{\alpha}_t \mid \boldsymbol{y}_t, \boldsymbol{\alpha}_{t+1}$ (in the artificial distribution $\widetilde{\mathrm{P}}$), and $\boldsymbol{\alpha}_t \mid \boldsymbol{y}_t, \boldsymbol{\alpha}_{t-1}, \boldsymbol{\alpha}_{t+1}$. However, if we assume that $g_t\left(\boldsymbol{y}_t \mid \boldsymbol{\alpha}_t\right)$ is log-concave in $\boldsymbol{\alpha}_t$, then it is easy to show that all of the previous three conditional distributions are unimodal. Hence, we can make a multivariate normal approximation as in Pitt and Shephard (1999). To do so, we make a second order Taylor expansion around some value $\boldsymbol{z}$ to arrive at

$$\begin{aligned}
k_t(\boldsymbol{\alpha}_t) &= \log g_t\left(\boldsymbol{y}_t \mid \boldsymbol{\eta}(\boldsymbol{\alpha}_t)\right), \qquad \boldsymbol{\eta}(\boldsymbol{\alpha}_t) = \mathbf{X}_t\boldsymbol{\alpha}_t + \boldsymbol{h}_t \\
\log g_t\left(\boldsymbol{y}_t \mid \boldsymbol{\alpha}_t\right) &\approx \mathrm{D}k_t(\boldsymbol{z})(\boldsymbol{\alpha}_t - \boldsymbol{z}) + \frac{1}{2}(\boldsymbol{\alpha}_t - \boldsymbol{z})^\top \mathrm{H}k_t(\boldsymbol{z})(\boldsymbol{\alpha}_t - \boldsymbol{z}) + \dots \\
&= \boldsymbol{\alpha}_t^\top \mathrm{D}k_t(\boldsymbol{z})^\top - \frac{1}{2}(\boldsymbol{\alpha}_t - \boldsymbol{z})^\top \left(-\mathrm{H}k_t(\boldsymbol{z})\right)(\boldsymbol{\alpha}_t - \boldsymbol{z}) + \dots \\
&= \boldsymbol{\alpha}_t^\top \left(-\mathrm{H}k_t(\boldsymbol{z})\right)\left(\boldsymbol{z} - \mathrm{H}k_t(\boldsymbol{z})^{-1}\mathrm{D}k_t(\boldsymbol{z})^\top\right) - \frac{1}{2}\boldsymbol{\alpha}_t^\top \left(-\mathrm{H}k_t(\boldsymbol{z})\right)\boldsymbol{\alpha}_t + \dots
\end{aligned}$$

where $\dots$ includes the zero order term, $\mathrm{D}k_t$ is the Jacobian, and $\mathrm{H}k_t$ denotes the Hessian. I add a subscript to D and H to indicate which variable the Jacobian or Hessian are with respect to. We find that

$$\begin{aligned}
\mathrm{H}k_t(\boldsymbol{z}) &= \mathrm{D}_{\boldsymbol{\alpha}_t}\boldsymbol{\eta}(\boldsymbol{z})^\top \mathrm{H}_{\boldsymbol{\eta}} \log g_t\left(\boldsymbol{y}_t \mid \boldsymbol{\eta}(\boldsymbol{z})\right)\mathrm{D}_{\boldsymbol{\alpha}_t}\boldsymbol{\eta}(\boldsymbol{z}) \\
&= \mathbf{X}_t^\top(-\mathbf{G}_t(\boldsymbol{z}))\mathbf{X}_t, \qquad \mathbf{G}_t(\boldsymbol{z}) = -\mathrm{H}_{\boldsymbol{\eta}} \log g_t\left(\boldsymbol{y}_t \mid \boldsymbol{\eta}(\boldsymbol{z})\right)
\end{aligned}$$

which follows from the chain rule and where we use that $\mathrm{H}_{\boldsymbol{\alpha}_t}\boldsymbol{\eta}(\boldsymbol{z}) = \mathbf{0}$. Thus,

$$\begin{aligned}
\log g_t\left(\boldsymbol{y}_t \mid \boldsymbol{\alpha}_t\right) &\approx \boldsymbol{\alpha}_t^\top \mathbf{X}_t^\top \mathbf{G}_t(\boldsymbol{z})\boldsymbol{u}_t(\boldsymbol{z}) - \frac{1}{2}\boldsymbol{\alpha}_t \mathbf{X}_t^\top \mathbf{G}_t(\boldsymbol{z})\mathbf{X}_t\boldsymbol{\alpha}_t \\
\boldsymbol{u}_t(\boldsymbol{z}) &= \mathbf{X}_t\boldsymbol{z} - \mathbf{X}_t\mathrm{H}k_t(\boldsymbol{z})^{-1}\mathrm{D}k_t(\boldsymbol{z})^\top
\end{aligned}$$

This yields the following multivariate normal approximation

$$g_t\left(\boldsymbol{y}_t \mid \boldsymbol{\alpha}_t\right) \approx \phi\left(\mathbf{X}_t\boldsymbol{\alpha}_t \,\middle|\, \boldsymbol{u}_t(\boldsymbol{z}), \mathbf{G}_t(\boldsymbol{z})^{-1}\right)$$

The Taylor approximation is easily used in the proposal distributions. For instance, for given $\boldsymbol{z}$, we arrive at the following mean and covariance matrix analogues to Equation (4.3) and (4.4) in the proposal distribution in the forward particle filter

$$\begin{aligned}
\boldsymbol{\Sigma}_t(\boldsymbol{z}) &= \left(\mathbf{Q}^{-1} + \mathbf{X}_t^\top \mathbf{G}_t(\boldsymbol{z})\mathbf{X}_t\right)^{-1} \\
\boldsymbol{\mu}_t(\boldsymbol{x}, \boldsymbol{z}) &= \boldsymbol{\Sigma}_t(\boldsymbol{z})\left(\mathbf{Q}^{-1}\mathbf{F}\boldsymbol{x} + \mathbf{X}_t^\top \mathbf{G}_t(\boldsymbol{z})\boldsymbol{u}_t(\boldsymbol{z})\right)
\end{aligned}$$

As for the resampling weights, we can use

$$\begin{aligned}
\widehat{\boldsymbol{\alpha}} &= \boldsymbol{\mu}_t(\boldsymbol{\alpha}_{t-1}^{(j)}, \boldsymbol{z}) \\
\beta_t^{(j)} &\propto p\left(\boldsymbol{y}_t \,\middle|\, \boldsymbol{\alpha}_{t-1}^{(j)}\right) w_{t-1}^{(j)} \\
&= \frac{p\left(\boldsymbol{y}_t \,\middle|\, \boldsymbol{\alpha}_{t-1}^{(j)}\right)}{g_t\left(\boldsymbol{y}_t \mid \widehat{\boldsymbol{\alpha}}\right) f\left(\widehat{\boldsymbol{\alpha}} \,\middle|\, \boldsymbol{\alpha}_{t-1}^{(j)}\right)} g_t\left(\boldsymbol{y}_t \mid \widehat{\boldsymbol{\alpha}}\right) f\left(\widehat{\boldsymbol{\alpha}} \,\middle|\, \boldsymbol{\alpha}_{t-1}^{(j)}\right) w_{t-1}^{(j)} \\
&= \frac{g_t\left(\boldsymbol{y}_t \mid \widehat{\boldsymbol{\alpha}}\right) f\left(\widehat{\boldsymbol{\alpha}} \,\middle|\, \boldsymbol{\alpha}_{t-1}^{(j)}\right) w_{t-1}^{(j)}}{p\left(\widehat{\boldsymbol{\alpha}} \,\middle|\, \boldsymbol{\alpha}_{t-1}^{(j)}, \boldsymbol{y}_t\right)} \\
&\approx \frac{g_t\left(\boldsymbol{y}_t \mid \widehat{\boldsymbol{\alpha}}\right) f\left(\widehat{\boldsymbol{\alpha}} \,\middle|\, \boldsymbol{\alpha}_{t-1}^{(j)}\right) w_{t-1}^{(j)}}{q\left(\widehat{\boldsymbol{\alpha}} \,\middle|\, \boldsymbol{\alpha}_{t-1}^{(j)}, \boldsymbol{y}_t\right)}
\end{aligned}$$

as in Fearnhead et al. (2010). We can approximate the backwards particle filter resampling weights in Equation (4.11) in a similar way

$$\begin{aligned}
\widetilde{\beta}_t^{(k)} &\propto \widetilde{p}\left(\boldsymbol{y}_t \,\middle|\, \widetilde{\boldsymbol{\alpha}}_{t+1}^{(k)}\right) \widetilde{w}_{t+1}^{(k)} \\
&\approx \frac{g_t\left(\boldsymbol{y}_t \mid \widehat{\boldsymbol{\alpha}}\right) \widetilde{p}\left(\widehat{\boldsymbol{\alpha}} \,\middle|\, \widetilde{\boldsymbol{\alpha}}_{t+1}^{(k)}\right) \widetilde{w}_{t+1}^{(k)}}{\widetilde{q}\left(\widehat{\boldsymbol{\alpha}} \,\middle|\, \boldsymbol{y}_t, \widetilde{\boldsymbol{\alpha}}_{t+1}^{(k)}\right)} \\
&= \frac{g_t\left(\boldsymbol{y}_t \mid \widehat{\boldsymbol{\alpha}}\right) f\left(\widetilde{\boldsymbol{\alpha}}_{t+1}^{(k)} \,\middle|\, \widehat{\boldsymbol{\alpha}}\right) \gamma_t(\widehat{\boldsymbol{\alpha}}) \widetilde{w}_{t+1}^{(k)}}{\widetilde{q}\left(\widehat{\boldsymbol{\alpha}} \,\middle|\, \boldsymbol{y}_t, \widetilde{\boldsymbol{\alpha}}_{t+1}^{(k)}\right) \gamma_{t+1}(\widetilde{\boldsymbol{\alpha}}_{t+1}^{(k)})} && (4.21) \\
\widehat{\boldsymbol{\alpha}} &= \overleftarrow{\boldsymbol{\mu}}(\widetilde{\boldsymbol{\alpha}}_{t+1}^{(k)}, \boldsymbol{z}) \\
\overleftarrow{\boldsymbol{\mu}}(\boldsymbol{x}, \boldsymbol{z}) &= \overleftarrow{\boldsymbol{\Sigma}}_t(\boldsymbol{z}) \left(\mathbf{P}_t^{-1}\boldsymbol{m}_t + \mathbf{F}^\top \mathbf{Q}^{-1}\boldsymbol{x} + \mathbf{X}_t^\top \mathbf{G}_t(\boldsymbol{z})\boldsymbol{u}_t(\boldsymbol{z})\right) && (4.22) \\
\overleftarrow{\boldsymbol{\Sigma}}_t(\boldsymbol{z}) &= \left(\mathbf{P}_t^{-1} + \mathbf{F}^\top \mathbf{Q}^{-1}\mathbf{F} + \mathbf{X}_t^\top \mathbf{G}_t(\boldsymbol{z})\mathbf{X}_t\right)^{-1} && (4.23)
\end{aligned}$$

We may also use a multivariate t-distribution for the proposal distribution to yield heavier tails than we do with the multivariate normal distribution. This may be important as too light tailed proposal distributions (relative to the target) can yield few large importance weights.

4.2.1 Where to Make the Expansion

An option is to make the Taylor expansion at a mode for each particle or particle pair in the smoothing step. This yields

$$\begin{aligned}
\boldsymbol{z}^{(j)} &= \arg\max_{\boldsymbol{z}} g_t\left(\boldsymbol{y}_t \mid \boldsymbol{z}\right) f\left(\boldsymbol{z} \,\middle|\, \boldsymbol{\alpha}_{t-1}^{(j)}\right) \\
\boldsymbol{z}^{(k)} &= \arg\max_{\boldsymbol{z}} g_t\left(\boldsymbol{y}_t \mid \boldsymbol{z}\right) \gamma_t(\boldsymbol{z}) f\left(\widetilde{\boldsymbol{\alpha}}_{t+1}^{(k)} \,\middle|\, \boldsymbol{z}\right) \\
\boldsymbol{z}^{(i)} &= \arg\max_{\boldsymbol{z}} f\left(\boldsymbol{z} \,\middle|\, \boldsymbol{\alpha}_{t-1}^{(j_i)}\right) g_t\left(\boldsymbol{y}_t \mid \boldsymbol{z}\right) f\left(\widetilde{\boldsymbol{\alpha}}_{t+1}^{(k_i)} \,\middle|\, \boldsymbol{z}\right)
\end{aligned}$$

for, respectively, the forward particle filter, the backward particle filter, and the smoother. The downside is a $\mathcal{O}\left(r^2 n_{\max} N_S\right)$ or $\mathcal{O}\left(r^2 n_{\max} N\right)$ computational complexity at each time step given that we have to evaluate $\mathbf{X}_t^\top \mathbf{G}_t(\boldsymbol{z})\mathbf{X}_t$ for each particle or particle pair. Instead, we can make the approximation once at each time step at, respectively, $\sum_{i=1}^N w_{t-1}^{(j)}\boldsymbol{\alpha}_{t-1}^{(j)}$, $\sum_{i=1}^N \widetilde{w}_{t+1}^{(j)}\widetilde{\boldsymbol{\alpha}}_{t+1}^{(j)}$, and

$$\left(\mathbf{Q}^{-1} + \mathbf{F}^\top\mathbf{Q}^{-1}\mathbf{F}\right)^{-1}\left(\mathbf{Q}^{-1}\mathbf{F}\sum_{i=1}^N w_{t-1}^{(j)}\boldsymbol{\alpha}_{t-1}^{(j)} + \mathbf{F}^\top\mathbf{Q}^{-1}\sum_{i=1}^N \widetilde{w}_{t+1}^{(j)}\widetilde{\boldsymbol{\alpha}}_{t+1}^{(j)}\right)$$

which will reduce the computational complexity at each time step to $\mathcal{O}\left(rn_{\max}N_S + r^2 n_{\max}\right)$ or $\mathcal{O}\left(rn_{\max}N + r^2 n_{\max}\right)$.

4.3 Log-Likelihood Evaluation and Parameter Estimation

In this section, I show an example of parameter estimation in the first-order random walk using a Monte Carlo EM-algorithm. Then I cover the general vector autoregression model and how one can estimate the fixed effects. See Del Moral et al. (2010), Kantas et al. (2015), Schön et al. (2011) for general discussion of parameter estimation with particle filters and smoothers. Firstly, though, I will remark that we can approximate the log-likelihood for a particular value of $\{\mathbf{Q}, \mathbf{Q}_0, \boldsymbol{\mu}_0, \mathbf{F}\}$ as described in Doucet and Johansen (2009) and Malik and Pitt (2011), using the forward particle filter shown in Algorithm 2. Details are omitted here for the sake of brevity.

The formulas for parameter estimation for the first-order random walk model are particularly simple. We need to estimate $\mathbf{Q}$ and $\boldsymbol{a}_0$ elements of $\boldsymbol{\varphi} = \{\mathbf{Q}, \mathbf{Q}_0, \boldsymbol{\mu}_0\}$. We do this by running Algorithm 1 for the current $\boldsymbol{\varphi}$. This yields the following quantities from the E-step

$$\begin{aligned}
\boldsymbol{t}_t^{(\boldsymbol{\varphi})} &\approx \sum_{i=1}^{N_s} \widehat{\boldsymbol{\alpha}}_t^{(i)}\widehat{w}_t^{(i)} \\
\boldsymbol{T}_t^{(\boldsymbol{\varphi})} &\approx \sum_{i=1}^{N_s}\left(\widehat{\boldsymbol{\alpha}}_t^{(i)} - \mathbf{F}\boldsymbol{\alpha}_{t-1}^{(j_{it})}\right)\left(\widehat{\boldsymbol{\alpha}}_t^{(i)} - \mathbf{F}\boldsymbol{\alpha}_{t-1}^{(j_{it})}\right)^\top \widehat{w}_t^{(i)}
\end{aligned} \tag{4.24}$$

where we have extended the notation in Algorithm 1 such that superscript j_{it} is the index from forward cloud at time $t-1$ matching with ith smoothed particle at time t. Then we carry out the M-step by updating $\boldsymbol{\mu}_0$ and $\mathbf{Q}$ given the summary statistics above

$$\boldsymbol{\mu}_0 = \boldsymbol{t}_1^{(\boldsymbol{\varphi})} \qquad \mathbf{Q} = \frac{1}{d}\sum_{t=1}^{d} \boldsymbol{T}_t^{(\boldsymbol{\varphi})} \tag{4.25}$$

We then take another iteration of the EM-algorithm with the new $\boldsymbol{\mu}_0$ and $\mathbf{Q}$ and repeat until a convergence criteria is satisfied.

4.3.1 Vector Autoregression Models

We start by defining the following matrices to cover estimation in general vector autoregression models for the latent space variable

$$\begin{aligned}
\mathbf{N} &= \left(\widehat{\boldsymbol{\alpha}}_2^{(1)}, \dots, \widehat{\boldsymbol{\alpha}}_2^{(N_s)}, \widehat{\boldsymbol{\alpha}}_3^{(1)}, \dots, \widehat{\boldsymbol{\alpha}}_3^{(N_s)}, \widehat{\boldsymbol{\alpha}}_4^{(1)}, \dots, \widehat{\boldsymbol{\alpha}}_d^{(N_s)}\right)^\top \\
\mathbf{M} &= \left(\boldsymbol{\alpha}_1^{(j_{12})}, \dots, \boldsymbol{\alpha}_1^{(j_{N_s 2})}, \boldsymbol{\alpha}_2^{(j_{13})}, \dots, \boldsymbol{\alpha}_2^{(j_{N_s 3})}, \boldsymbol{\alpha}_3^{(j_{14})}, \dots, \boldsymbol{\alpha}_{d-1}^{(j_{N_s d})}\right)^\top \\
\mathbf{W} &= \operatorname{diag}\left(\widehat{w}_2^{(1)}, \dots, \widehat{w}_2^{(N_s)}, \widehat{w}_3^{(1)}, \dots, \widehat{w}_3^{(N_s)}, \widehat{w}_4^{(1)}, \dots, \widehat{w}_d^{(N_s)}\right)
\end{aligned}$$

where diag$(\cdot)$ is a diagonal matrix. We suppress the dependence above on the result of the E-step in a given iteration of the EM-algorithm to ease the notation. The goal is to estimate $\mathbf{F}$ and $\mathbf{Q}$ in Equation (4.1). We can find that the M-step maximizers are

$$\widehat{\mathbf{F}}^\top = \left(\mathbf{M}^\top\mathbf{W}\mathbf{M}\right)^{-1}\mathbf{M}^\top\mathbf{W}\mathbf{N} \tag{4.26}$$

$$\widehat{\mathbf{Q}} = \frac{1}{d-1}\left(\mathbf{N} - \widehat{\mathbf{F}}\mathbf{M}\right)^\top \mathbf{W}\left(\mathbf{N} - \widehat{\mathbf{F}}\mathbf{M}\right) \tag{4.27}$$

which are the typical vector autoregression estimators with weights. Equation (4.26) and (4.27) can easily be computed in parallel using a QR decomposition as in the `bam` function in the **mgcv** package with a low memory footprint (see Wood et al., 2015). This is currently implemented. However, the gains from a parallel implementation may be small, because the computational complexity is independent of the number of observations. The computation involved here is often fast relative to other parts of the Monte Carlo EM-algorithm because the dimension of the state vector is relatively small.

4.3.2 Restricted Vector Autoregression Models

Suppose that we want to restrict some of the parameters of $\mathbf{F}$ and $\mathbf{Q}$. Let

$$\begin{aligned}
(s_1, s_2, \dots, s_r)^\top &= \mathbf{J}\boldsymbol{\psi} \\
(o_{21}, o_{31}, \dots, o_{r1}, o_{32}, \dots, o_{r2}, o_{43}, \dots, o_{r,r-1})^\top &= \mathbf{K}\boldsymbol{\phi}
\end{aligned}$$

Then we can restrict the model such that

$$\begin{aligned}
\sigma_i &= \exp(s_i) & \rho_{ij} &= \frac{2}{1+\exp(-o_{ij})} - 1 \\
\operatorname{vec}(\mathbf{F}) &= \mathbf{G}\boldsymbol{\theta} & \mathbf{Q} &= \mathbf{V}\mathbf{C}\mathbf{V} \\
\mathbf{V} &= \begin{pmatrix} \sigma_1 & 0 & \cdots & 0 \\ 0 & \sigma_2 & \ddots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \dots & 0 & \sigma_r \end{pmatrix} & \mathbf{C} &= \begin{pmatrix} 1 & \rho_{21} & \cdots & \rho_{r1} \\ \rho_{21} & 1 & \ddots & \rho_{r2} \\ \vdots & \ddots & \ddots & \vdots \\ \rho_{r1} & \dots & \rho_{r,r-1} & 1 \end{pmatrix}
\end{aligned}$$

and where $\operatorname{vec}(\cdot)$ is the vectorization function which stacks the the columns of a matrix from left to right. For example,

$$\begin{aligned}
\operatorname{vec}(\mathbf{A}) &= (a_{11}, a_{21}, a_{31}, a_{12}, a_{22}, a_{32}, a_{13}, a_{23}, a_{33})^\top \\
\mathbf{A} &= \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix}
\end{aligned}$$

$\mathbf{G} \in \mathbb{R}^{r^2 \times g}$ is a known matrix with $g \leq r^2$, and we assume that it has full column rank. Similarly, $\mathbf{J} \in \mathbb{R}^{r \times l}$ with $l \leq r$ and $\mathbf{K} \in \mathbb{R}^{r(r-1)/2 \times k}$ with $k \leq r(r-1)/2$ are known and have full column rank. We assume that $\mathbf{G}$ is such that $\mathbf{F}$ is nonsingular for some $\boldsymbol{\theta}$ because Equation (4.10) is used. Further, we assume that $\mathbf{J}$ and $\mathbf{K}$ are such that $\mathbf{Q}$ is a positive definite matrix for some $\boldsymbol{\psi}$ and $\boldsymbol{\phi}$ pair. $\mathbf{V}$ is a diagonal matrix containing the standard deviations, and $\mathbf{C}$ is the correlation matrix.

We cannot jointly maximize $\boldsymbol{\theta}$, $\boldsymbol{\psi}$, and $\boldsymbol{\phi}$ analytically, but we can maximize $\boldsymbol{\theta}$ analytically conditional on $\boldsymbol{\psi}$ and $\boldsymbol{\phi}$. Hence, we can employ a Monte Carlo expectation conditional maximization algorithm in which we take two so-called conditional maximization steps (see Meng and Rubin, 1993, for the non-Monte Carlo expectation maximization algorithm). The first conditional maximization step is

$$\boldsymbol{\theta}^{(i+1)} = \mathbf{G}^+ \left(\mathbf{Q}^{(i)} \otimes \left(\mathbf{M}^\top \mathbf{W} \mathbf{M}\right)^{-1}\right) \mathbf{G}^{+\top} \mathbf{G}^\top \operatorname{vec}\left(\mathbf{M}^\top \mathbf{W} \mathbf{N} \mathbf{Q}^{-(i)}\right) \tag{4.28}$$

where $\otimes$ is the Kronecker product and $\mathbf{Q}^{-(i)}$ is the inverse of $\mathbf{Q}^{(i)}$ and $\mathbf{G}^+$ is a pseudoinverse of $\mathbf{G}$. Equation (4.28) is easily computed with the QR decomposition we make to compute for Equation (4.26). Having obtained the new $\boldsymbol{\theta}^{(i+1)}$, we update $\mathbf{F}$ and denote the new estimate $\widehat{\mathbf{F}}^{(i+1)}$. The second conditional maximization step where we update $\boldsymbol{\psi}$ and $\boldsymbol{\phi}$ is

$$\begin{aligned}
\mathbf{Z} &= \left(\mathbf{N} - \widehat{\mathbf{F}}^{(i+1)}\mathbf{M}\right)^\top \mathbf{W} \left(\mathbf{N} - \widehat{\mathbf{F}}^{(i+1)}\mathbf{M}\right) \\
\boldsymbol{\psi}^{(i+1)}, \boldsymbol{\phi}^{(i+1)} &= \operatorname*{arg\,max}_{\boldsymbol{\psi},\boldsymbol{\phi}} -(d-1)\log\left|\mathbf{Q}(\boldsymbol{\psi},\boldsymbol{\phi})\right| - \operatorname{tr}\left(\mathbf{Q}(\boldsymbol{\psi},\boldsymbol{\phi})^{-1}\mathbf{Z}\right)
\end{aligned}$$

which can be done numerically. We have made $\mathbf{Q}$'s dependence on $\boldsymbol{\psi}$ and $\boldsymbol{\phi}$ explicit to emphasize which factors are affected. $\mathbf{C}$ may not be a valid correlation matrix for all $\boldsymbol{\phi} \in \mathbb{R}^k$ for some choices of $\mathbf{K}$. Thus, the numerical optimization algorithm is constrained to valid correlation matrices. This completes the two conditional maximization steps. The next E-step is then performed using $\boldsymbol{\theta}^{(i+1)}$, $\boldsymbol{\psi}^{(i+1)}$, $\boldsymbol{\phi}^{(i+1)}$. Meng and Rubin (1993, see the discussion) comments that it may be beneficial to perform an E-step between each conditional maximization step when the E-step is relatively cheap. This is not the case here because all the above computations are independent of the number observations, $n_{\max}$. Thus, if we have a moderately large number of observations at each time point relative to the dimension of the state vector, then the E-step will use most of the computation time.

4.3.3 Estimating Fixed Effect Coefficients

Next, we turn to estimating the fixed effect coefficients, $\boldsymbol{\omega}$, in Equation (4.1). If we assume that observations, y_{it}s, are from an exponential family conditional on the state vector and covariates (or we can transform the data into an exponential family that differ only by a normalization constant), then it is easy to show that the M-step estimator can be found as the MLE to a specific generalized linear model with N_s observations for each y_{it}, differing only by an offset term and a weight. The offset term comes from the $\boldsymbol{x}_{it}^\top \widehat{\boldsymbol{\alpha}}_j^{(t)}$ term in Equation (4.1) for each of the $j = 1, \dots, N_s$ smoothed particles. The corresponding weights are the smoothed weights, $\widehat{w}_j^{(t)}$. The problem can be solved in parallel using a QR decomposition as in Section 4.3.1. This is what is done in the current implementation. Currently, only one iteration of the iteratively re-weighted least squares is performed at each M-step due to (somewhat limited) empirical evidence that the fixed coefficients typically do not change much with further iteration in the M-step.

4.4 Other Filter and Smoother Options

There is a long list of other particle-based methods besides the implemented smoother from Fearnhead et al. (2010). This section will give a brief overview of some the options and argue for the alternative smoother that is implemented in the package. The $\mathcal{O}\left(N^2\right)$ two-filter smoother in Fearnhead et al. (2010) is going to be computationally expensive, because an approximation will be needed for Equation (8) in their article. The non-auxiliary version in Briers et al. (2009) is more feasible as it only requires evaluation of f in the smoothing part of the generalized two-filter smoother (see Equation (46) in their paper). Similar conclusions apply to the forward smoother in Del Moral et al. (2010) and the backward smoother as presented in Kantas et al. (2015). Both have a $\mathcal{O}\left(N^2\right)$ computational cost.

Despite the $\mathcal{O}\left(N^2\right)$ cost of the methods in Briers et al. (2009) and Del Moral et al. (2010), they may still be useful, because the computational cost in the smoothing step is independent of the number of observations, $n_{\max}$. Further, the computational cost can be reduced to a $\mathcal{O}\left(N \log(N)\right)$ average-case complexity with approximations as in Klaas et al. (2006). The method in Malik and Pitt (2011) can also be used to do continuous likelihood approximations as a function of the unknown parameters.

Kantas et al. (2015) show, empirically that only using a forward filter may be an effective method. However, the example is with an univariate outcome ($n_{\max} = 1$, not to be confused with the number of periods d). In the problems shown in this chapter, the computational complexity of the forward filter is at least $\mathcal{O}\left(dNn_{\max}r\right)$. Every new particle yields an $\mathcal{O}\left(dn_{\max}r\right)$ cost, which is expensive due to the large number of observed outcomes, $n_{\max}$. Thus, the considerations are different and a $\mathcal{O}\left(dNn_{\max}r + N^2\right)$ method will not make a big difference unless N is large. Thus, one of the $\mathcal{O}\left(N^2\right)$ particle smoothers is also implemented in the **dynamichazard** package.

4.5 Generalized Two-Filter Smoother

The $\mathcal{O}\left(N^2\right)$ smoother from Briers et al. (2009) is also implemented because it is feasible for a moderate number of particles. Algorithm 4 shows this smoother. The weights in Equation (4.31) comes from the generalized two-filter formula. The arguments for the smoother is that

$$p\left(\boldsymbol{y}_{t:d} \mid \boldsymbol{\alpha}_t\right) = \widetilde{p}\left(\boldsymbol{y}_{t:d}\right)\frac{\widetilde{p}\left(\boldsymbol{\alpha}_t \mid \boldsymbol{y}_{t:d}\right)}{\gamma_t(\boldsymbol{\alpha}_t)}$$

which we can use to generalize the two-filter formula from Kitagawa (1994) as follows

$$\begin{aligned}
p\left(\boldsymbol{\alpha}_t \mid \boldsymbol{y}_{1:d}\right) &= \frac{p\left(\boldsymbol{\alpha}_t \mid \boldsymbol{y}_{1:t-1}\right)p\left(\boldsymbol{y}_{t:d} \mid \boldsymbol{\alpha}_t\right)}{p\left(\boldsymbol{y}_{t:d} \mid \boldsymbol{y}_{1:t-1}\right)} \\
&\propto p\left(\boldsymbol{\alpha}_t \mid \boldsymbol{y}_{1:t-1}\right)\widetilde{p}\left(\boldsymbol{y}_{t:d}\right)\frac{\widetilde{p}\left(\boldsymbol{\alpha}_t \mid \boldsymbol{y}_{t:d}\right)}{\gamma_t(\boldsymbol{\alpha}_t)} \\
&\propto \widetilde{p}\left(\boldsymbol{\alpha}_t \mid \boldsymbol{y}_{t:d}\right)\frac{\left[\int p\left(\boldsymbol{\alpha}_{t-1} \mid \boldsymbol{y}_{1:t-1}\right)f\left(\boldsymbol{\alpha}_t \mid \boldsymbol{\alpha}_{t-1}\right)\mathrm{d}\boldsymbol{\alpha}_{t-1}\right]}{\gamma_t(\boldsymbol{\alpha}_t)} \\
&\underset{\sim}{\propto} \sum_{i=1}^N \widetilde{w}_t^{(i)}\delta_{\widetilde{\boldsymbol{\alpha}}_t^{(i)}}\left(\boldsymbol{\alpha}_t\right)\frac{\left[\sum_{j=1}^N w_{t-1}^{(j)} f\left(\widetilde{\boldsymbol{\alpha}}_t^{(i)} \,\middle|\, \boldsymbol{\alpha}_{t-1}^{(j)}\right)\right]}{\gamma_t\left(\widetilde{\boldsymbol{\alpha}}_t^{(i)}\right)}
\end{aligned} \tag{4.29}$$

where $\underset{\sim}{\propto}$ means approximately proportional. Similar arguments lead to

$$\begin{aligned}
p\left(\boldsymbol{\alpha}_{t-1:t} \mid \boldsymbol{y}_{1:d}\right) &\propto p\left(\boldsymbol{\alpha}_{t-1:t} \mid \boldsymbol{y}_{1:t-1}\right)p\left(\boldsymbol{y}_{t:d} \mid \boldsymbol{\alpha}_{t-1:t}\right) \\
&\propto f\left(\boldsymbol{\alpha}_t \mid \boldsymbol{\alpha}_{t-1}\right)p\left(\boldsymbol{\alpha}_{t-1} \mid \boldsymbol{y}_{1:t-1}\right)\frac{\widetilde{p}\left(\boldsymbol{\alpha}_t \mid \boldsymbol{y}_{t:d}\right)}{\gamma_t(\boldsymbol{\alpha}_t)} \\
&\underset{\sim}{\propto} \sum_{i=1}^N\sum_{j=1}^N \widetilde{w}_t^{(i)}\delta_{\widetilde{\boldsymbol{\alpha}}_t^{(i)}}\left(\boldsymbol{\alpha}_t\right)\frac{\left[\sum_{k=1}^N w_{t-1}^{(k)} f\left(\widetilde{\boldsymbol{\alpha}}_t^{(i)} \,\middle|\, \boldsymbol{\alpha}_{t-1}^{(k)}\right)\right]}{\gamma_t\left(\widetilde{\boldsymbol{\alpha}}_t^{(i)}\right)} \\
&\qquad\qquad \cdot \frac{w_{t-1}^{(j)}\delta_{\boldsymbol{\alpha}_{t-1}^{(j)}}\left(\boldsymbol{\alpha}_{t-1}\right)f\left(\widetilde{\boldsymbol{\alpha}}_t^{(i)} \,\middle|\, \boldsymbol{\alpha}_{t-1}^{(j)}\right)}{\left[\sum_{k=1}^N w_{t-1}^{(k)} f\left(\widetilde{\boldsymbol{\alpha}}_t^{(i)} \,\middle|\, \boldsymbol{\alpha}_{t-1}^{(k)}\right)\right]} \\
&\propto \sum_{i=1}^N\sum_{j=1}^N \widehat{w}_t^{(i,j)}\delta_{\widetilde{\boldsymbol{\alpha}}_t^{(i)}}\left(\boldsymbol{\alpha}_t\right)\delta_{\boldsymbol{\alpha}_{t-1}^{(j)}}\left(\boldsymbol{\alpha}_{t-1}\right)
\end{aligned}$$

where

$$\widehat{w}_t^{(i,j)} = \widehat{w}_t^{(i)} \frac{w_{t-1}^{(j)} f\left(\widetilde{\boldsymbol{\alpha}}_t^{(i)} \,\middle|\, \boldsymbol{\alpha}_{t-1}^{(j)} \right)}{\left[\sum_{j=1}^N w_{t-1}^{(j)} f\left(\widetilde{\boldsymbol{\alpha}}_t^{(i)} \,\middle|\, \boldsymbol{\alpha}_{t-1}^{(j)} \right) \right]} \tag{4.30}$$

We need the latter for the Monte Carlo EM-algorithm.

Algorithm 4 $\mathcal{O}\left(N^2\right)$ generalized two-filter smoother suggested by Briers et al. (2009).

```
Input:
    $\mathbf{Q}, \mathbf{Q}_0, \boldsymbol{a}_0, \mathbf{X}_1, \dots, \mathbf{X}_d, \boldsymbol{y}_1, \dots, \boldsymbol{y}_d, R_1, \dots, R_d, \boldsymbol{\omega}$
procedure Filter forward
    Run a forward particle filter to yield particle clouds $\left\{ \boldsymbol{\alpha}_t^{(j)}, w_t^{(j)}, \beta_{t+1}^{(j)} \right\}_{j=1,\dots,N}$ approximating $p(\boldsymbol{\alpha}_t \mid \boldsymbol{y}_{1:t})$ for $t = 0, 1, \dots, d$. See Algorithm 2.
procedure Filter backwards
    Run a similar backward filter to yield $\left\{ \widetilde{\boldsymbol{\alpha}}_t^{(k)}, \widetilde{w}_t^{(k)}, \widetilde{\beta}_{t-1}^{(k)} \right\}_{k=1,\dots,N}$ approximating $\widetilde{p}(\boldsymbol{\alpha}_t \mid \boldsymbol{y}_{t:d})$ for $t = d+1, d, d-1, \dots, 1$. See Algorithm 3.
procedure Smooth (combine)
    for $t = 1, \dots, d$ do
        Assign each backward filter particle a smoothing weight given by
            $$\widehat{w}_t^{(i)} \propto \widetilde{w}_t^{(i)} \frac{\left[ \sum_{j=1}^N w_{t-1}^{(j)} f\left( \widetilde{\boldsymbol{\alpha}}_t^{(i)} \,\middle|\, \boldsymbol{\alpha}_{t-1}^{(j)} \right) \right]}{\gamma_t\left( \widetilde{\boldsymbol{\alpha}}_t^{(i)} \right)} \tag{4.31}$$
```

With the results above, we can show the arguments behind the smoother from Fearnhead et al. (2010). Similar to Equation (4.29), we find that

$$\begin{aligned}
p(\boldsymbol{\alpha}_t \mid \boldsymbol{y}_{1:d}) &\propto p(\boldsymbol{\alpha}_t \mid \boldsymbol{y}_{1:t-1})\, p(\boldsymbol{y}_{t:d} \mid \boldsymbol{\alpha}_t) \\
&= p(\boldsymbol{\alpha}_t \mid \boldsymbol{y}_{1:t-1})\, g_t(\boldsymbol{y}_t \mid \boldsymbol{\alpha}_t)\, p(\boldsymbol{y}_{t+1:d} \mid \boldsymbol{\alpha}_t) \\
&= \int f(\boldsymbol{\alpha}_t \mid \boldsymbol{\alpha}_{t-1})\, p(\boldsymbol{\alpha}_{t-1} \mid \boldsymbol{y}_{1:t-1})\, \mathrm{d}\boldsymbol{\alpha}_{t-1} g_t(\boldsymbol{y}_t \mid \boldsymbol{\alpha}_t) \\
&\qquad\qquad \cdot \int f(\boldsymbol{\alpha}_{t+1} \mid \boldsymbol{\alpha}_t)\, p(\boldsymbol{y}_{t+1:d} \mid \boldsymbol{\alpha}_{t+1})\, \mathrm{d}\boldsymbol{\alpha}_{t+1} \\
&\propto \int f(\boldsymbol{\alpha}_t \mid \boldsymbol{\alpha}_{t-1})\, p(\boldsymbol{\alpha}_{t-1} \mid \boldsymbol{y}_{1:t-1})\, \mathrm{d}\boldsymbol{\alpha}_{t-1} g_t(\boldsymbol{y}_t \mid \boldsymbol{\alpha}_t) \\
&\qquad\qquad \cdot \int f(\boldsymbol{\alpha}_{t+1} \mid \boldsymbol{\alpha}_t) \frac{\widetilde{p}(\boldsymbol{\alpha}_{t+1} \mid \boldsymbol{y}_{t+1:d})}{\gamma_{t+1}(\boldsymbol{\alpha}_{t+1})}\, \mathrm{d}\boldsymbol{\alpha}_{t+1} \\
&\overset{\propto}{\sim} \sum_{j=1}^N \sum_{k=1}^N f\left(\boldsymbol{\alpha}_t \,\middle|\, \boldsymbol{\alpha}_{t-1}^{(j)} \right) w_{t-1}^{(j)} g_t(\boldsymbol{y}_t \mid \boldsymbol{\alpha}_t)\, f\left(\widetilde{\boldsymbol{\alpha}}_{t+1}^{(k)} \,\middle|\, \boldsymbol{\alpha}_t \right) \frac{\widetilde{w}_{t+1}^{(k)}}{\gamma_{t+1}(\widetilde{\boldsymbol{\alpha}}_{t+1}^{(k)})}
\end{aligned}$$

Thus, we can sample $\boldsymbol{\alpha}_t$ from a proposal distribution, given the time $t-1$ forward filter particle, $\boldsymbol{\alpha}_{t-1}^{(j)}$, and time $t+1$ backward filter particle, $\widetilde{\boldsymbol{\alpha}}_{t+1}^{(k)}$, for all N^2 particle pairs. Alternatively, we can sample the $t-1$ and $t+1$ particles independently, which yield Algorithm 1. Further, we can find that

$$\begin{aligned}
p\left(\boldsymbol{\alpha}_{t-1:t} \mid \boldsymbol{y}_{1:d}\right) &= p\left(\boldsymbol{\alpha}_{t-1:t} \mid \boldsymbol{y}_{1:t-1}\right) g_t\left(\boldsymbol{y}_t \mid \boldsymbol{\alpha}_{t-1:t}\right) p\left(\boldsymbol{y}_{t+1:d} \mid \boldsymbol{\alpha}_{t-1:t}\right) \\
&\propto f\left(\boldsymbol{\alpha}_t \mid \boldsymbol{\alpha}_{t-1}\right) p\left(\boldsymbol{\alpha}_{t-1} \mid \boldsymbol{y}_{1:t-1}\right) g_t\left(\boldsymbol{y}_t \mid \boldsymbol{\alpha}_t\right) \\
&\qquad\qquad \cdot \int f\left(\boldsymbol{\alpha}_{t+1} \mid \boldsymbol{\alpha}_t\right) \frac{\widetilde{p}\left(\boldsymbol{\alpha}_{t+1} \mid \boldsymbol{y}_{t+1:d}\right)}{\gamma_{t+1}(\boldsymbol{\alpha}_{t+1})} \mathrm{d}\boldsymbol{\alpha}_{t+1} \\
&\overset{\propto}{\sim} \sum_{i=1}^{N_s} \delta_{\widehat{\boldsymbol{\alpha}}_t^{(i)}}\left(\boldsymbol{\alpha}_t\right) \delta_{\boldsymbol{\alpha}_{t-1}^{(j_i)}}\left(\boldsymbol{\alpha}_t\right) f\left(\widehat{\boldsymbol{\alpha}}_t^{(i)} \;\middle|\; \boldsymbol{\alpha}_{t-1}^{(j_i)}\right) w_{t-1}^{(j_i)} g_t\left(\boldsymbol{y}_t \;\middle|\; \widehat{\boldsymbol{\alpha}}_t^{(i)}\right) \\
&\qquad\qquad \cdot \int f\left(\boldsymbol{\alpha}_{t+1} \;\middle|\; \widehat{\boldsymbol{\alpha}}_t^{(i)}\right) \frac{\widetilde{p}\left(\boldsymbol{\alpha}_{t+1} \mid \boldsymbol{y}_{t+1:d}\right)}{\gamma_{t+1}(\boldsymbol{\alpha}_{t+1})} \mathrm{d}\boldsymbol{\alpha}_{t+1} \\
&\overset{\propto}{\sim} \sum_{i=1}^{N_s} \widehat{w}_t^{(i)} \delta_{\widehat{\boldsymbol{\alpha}}_t^{(i)}}\left(\boldsymbol{\alpha}_t\right) \delta_{\boldsymbol{\alpha}_{t-1}^{(j_i)}}\left(\boldsymbol{\alpha}_t\right)
\end{aligned}$$

where superscripts j_i are used as in Algorithm 1 implicitly dependent on t.

4.6 Gradient and Observed Information Matrix

An alternative to the Monte Carlo EM-algorithm is to approximate the gradient and use it to perform the maximization with a gradient descent algorithm. Moreover, one may be interested in the observed information matrix e.g., to get approximate standard errors for the coefficient estimates. Two methods are implemented to make such approximations. The first method is the method covered in Cappé et al. (2005, section 8.3 and chapter 11). Its advantage is that it uses the output from the forward particle filter. However, the variance of the estimates increase at least quadratically in time, d (Poyiadjis et al., 2011). An alternative is to use the method shown by Poyiadjis et al. (2011). Like the smoothing algorithm from Briers et al. (2009), this method has the disadvantage of having a computational complexity that is quadratic in the number of particles, N.

I will give a brief introduction to the two methods in this section. What is presented here closely follows Poyiadjis et al. (2011). First, we will need some notation. We denote the complete data log-likelihood by

$$\begin{aligned}
c\left(\boldsymbol{y}_{1:t}, \boldsymbol{\alpha}_{0:t}\right) &= \log h\left(\boldsymbol{y}_{1:t}, \boldsymbol{\alpha}_{0:t}\right) \\
h\left(\boldsymbol{y}_{1:t}, \boldsymbol{\alpha}_{0:t}\right) &= \nu\left(\boldsymbol{\alpha}_0\right) \prod_{k=1}^{t} g_k\left(\boldsymbol{y}_k \mid \boldsymbol{\alpha}_k\right) f\left(\boldsymbol{\alpha}_k \mid \boldsymbol{\alpha}_{k-1}\right)
\end{aligned}$$

where ν is the density function of the state vector at time zero, all functions may implicitly depend on the unknown parameters, and the dimension of the arguments for c and h is given by the superscript of the arguments. A direct application of the results from Louis (1982) shows that the gradient of the observed data log-likelihood

$$o\left(\boldsymbol{y}_{1:t}\right) = \log \int h\left(\boldsymbol{y}_{1:t}, \boldsymbol{a}_{0:t}\right) \mathrm{d}\boldsymbol{a}_{0:t}$$

with respect to the unknown parameters are

$$\begin{aligned}\nabla o\left(\boldsymbol{y}_{1:t}\right) &= \frac{\partial}{\partial \boldsymbol{\theta}} \log \int h\left(\boldsymbol{y}_{1:t}, \boldsymbol{a}_{0:t}\right) \mathrm{d}\boldsymbol{a}_{0:t} = \frac{\int h'\left(\boldsymbol{y}_{1:t}, \boldsymbol{a}_{0:t}\right) \mathrm{d}\boldsymbol{a}_{0:t}}{\int h\left(\boldsymbol{y}_{1:t}, \boldsymbol{a}_{0:t}\right) \mathrm{d}\boldsymbol{a}_{0:t}} \\ &= \int c'\left(\boldsymbol{y}_{1:t}, \boldsymbol{a}_{0:t}\right) p\left(\, \boldsymbol{a}_{0:t} \mid \boldsymbol{y}_{1:t} \,\right) \mathrm{d}\boldsymbol{a}_{0:t}\end{aligned} \tag{4.32}$$

where $\boldsymbol{\theta}$ are the unknown parameters in the model, derivatives are with respect to $\boldsymbol{\theta}$, and $p\left(\, \boldsymbol{a}_{0:t} \mid \boldsymbol{y}_{1:t} \,\right)$ is the conditional density function of $\boldsymbol{a}_{0:t}$ given $\boldsymbol{y}_{1:t}$. Moreover, the Hessian is

$$\begin{aligned}\nabla^2 o\left(\boldsymbol{y}_{1:t}\right) &= \frac{\int h''\left(\boldsymbol{y}_{1:t}, \boldsymbol{a}_{0:t}\right) \mathrm{d}\boldsymbol{a}_{0:t}}{\int h\left(\boldsymbol{y}_{1:t}, \boldsymbol{a}_{0:t}\right) \mathrm{d}\boldsymbol{a}_{0:t}} - \nabla o\left(\boldsymbol{y}_{1:t}\right) \nabla o\left(\boldsymbol{y}_{1:t}\right)^\top \\ &= \int c''\left(\boldsymbol{y}_{1:t}, \boldsymbol{a}_{0:t}\right) p\left(\, \boldsymbol{a}_{0:t} \mid \boldsymbol{y}_{1:t} \,\right) \mathrm{d}\boldsymbol{a}_{0:t} \\ &\quad + \int c'\left(\boldsymbol{y}_{1:t}, \boldsymbol{a}_{0:t}\right) c'\left(\boldsymbol{y}_{1:t}, \boldsymbol{a}_{0:t}\right)^\top p\left(\, \boldsymbol{a}_{0:t} \mid \boldsymbol{y}_{1:t} \,\right) \mathrm{d}\boldsymbol{a}_{0:t} - \nabla o\left(\boldsymbol{y}_{1:t}\right) \nabla o\left(\boldsymbol{y}_{1:t}\right)^\top\end{aligned} \tag{4.33}$$

We can use that the forward particle filter yields not just an approximation of $p\left(\, \boldsymbol{a}_d \mid \boldsymbol{y}_{1:d} \,\right)$ but the entire path $p\left(\, \boldsymbol{a}_{0:d} \mid \boldsymbol{y}_{1:d} \,\right)$. That is, we can use the weights at time d from Equation (4.18) to make a discrete approximation of Equation (4.32) and (4.33) as shown in Cappé et al. (2005). However, the variance of the estimates grows at least quadratically in d. The issue is that for larger d, then few if not only one unique value of the initial state vector values ($\boldsymbol{\alpha}_i$ with $0 \leq i \ll d$) are present in the discrete approximation.

As an alternative, Poyiadjis et al. (2011) develop a marginal version of Equation (4.32) and (4.33). That is,

$$\begin{aligned}\tilde{c}\left(\boldsymbol{y}_{1:t}, \boldsymbol{\alpha}_t\right) &= \log \tilde{h}\left(\boldsymbol{y}_{1:t}, \boldsymbol{\alpha}_t\right) \\ \tilde{h}\left(\boldsymbol{y}_{1:t}, \boldsymbol{\alpha}_t\right) &= \begin{cases} g_t\left(\, \boldsymbol{y}_t \mid \boldsymbol{\alpha}_t \,\right) \int f\left(\, \boldsymbol{\alpha}_t \mid \boldsymbol{a}_{t-1} \,\right) \tilde{h}\left(\boldsymbol{y}_{1:(t-1)}, \boldsymbol{a}_{t-1}\right) \mathrm{d}\boldsymbol{a}_{t-1} & t > 0 \\ \nu\left(\boldsymbol{a}_t\right) & t = 0 \end{cases}\end{aligned} \tag{4.34}$$

$$\nabla o\left(\boldsymbol{y}_{1:t}\right) = \int \tilde{c}'\left(\boldsymbol{y}_{1:t}, \boldsymbol{a}_t\right) p\left(\, \boldsymbol{a}_t \mid \boldsymbol{y}_{1:t} \,\right) \mathrm{d}\boldsymbol{a}_t \tag{4.35}$$

$$\begin{aligned}\nabla^2 o\left(\boldsymbol{y}_{1:t}\right) &= \int \tilde{c}''\left(\boldsymbol{y}_{1:t}, \boldsymbol{a}_t\right) p\left(\, \boldsymbol{a}_t \mid \boldsymbol{y}_{1:t} \,\right) \mathrm{d}\boldsymbol{a}_t \\ &\qquad + \int \tilde{c}'\left(\boldsymbol{y}_{1:t}, \boldsymbol{a}_t\right) \tilde{c}'\left(\boldsymbol{y}_{1:t}, \boldsymbol{a}_t\right)^\top p\left(\, \boldsymbol{a}_t \mid \boldsymbol{y}_{1:t} \,\right) \mathrm{d}\boldsymbol{a}_t - \nabla o\left(\boldsymbol{y}_{1:t}\right) \nabla o\left(\boldsymbol{y}_{1:t}\right)^\top \\ &= \int \left(\frac{\tilde{h}''\left(\boldsymbol{y}_{1:t}, \boldsymbol{a}_t\right)}{\tilde{h}\left(\boldsymbol{y}_{1:t}, \boldsymbol{a}_t\right)} - \tilde{c}'\left(\boldsymbol{y}_{1:t}, \boldsymbol{a}_t\right) \tilde{c}'\left(\boldsymbol{y}_{1:t}, \boldsymbol{a}_t\right)^\top\right) p\left(\, \boldsymbol{a}_t \mid \boldsymbol{y}_{1:t} \,\right) \mathrm{d}\boldsymbol{a}_t \\ &\qquad + \int \tilde{c}'\left(\boldsymbol{y}_{1:t}, \boldsymbol{a}_t\right) \tilde{c}'\left(\boldsymbol{y}_{1:t}, \boldsymbol{a}_t\right)^\top p\left(\, \boldsymbol{a}_t \mid \boldsymbol{y}_{1:t} \,\right) \mathrm{d}\boldsymbol{a}_t - \nabla o\left(\boldsymbol{y}_{1:t}\right) \nabla o\left(\boldsymbol{y}_{1:t}\right)^\top\end{aligned} \tag{4.36}$$

While there is no analytical expression for the derivatives, one can establish a pointwise approximation recursively for $c'\left(\boldsymbol{y}_{1:t}, \boldsymbol{a}_t\right)$ and $c''\left(\boldsymbol{y}_{1:t}, \boldsymbol{a}_t\right)$, as suggested by Poyiadjis et al. (2011). To see this, let

$$s_t\left(\boldsymbol{\alpha}_t, \boldsymbol{\alpha}_{t-1}\right) = \log g_t\left(\, \boldsymbol{y}_t \mid \boldsymbol{\alpha}_t \,\right) + \log f\left(\, \boldsymbol{\alpha}_t \mid \boldsymbol{\alpha}_{t-1} \,\right)$$

Then

$$\tilde{h}'\left(\boldsymbol{y}_{1:t}, \boldsymbol{\alpha}_t\right) = \exp\left(o\left(\boldsymbol{y}_{1:t-1}\right)\right) g_t\left(\boldsymbol{y}_t \mid \boldsymbol{\alpha}_t\right) \int f\left(\boldsymbol{\alpha}_t \mid \boldsymbol{a}_{t-1}\right) p\left(\boldsymbol{a}_{t-1} \mid \boldsymbol{y}_{1:t-1}\right) \cdot \left(s_t'\left(\boldsymbol{\alpha}_t, \boldsymbol{a}_{t-1}\right) + \tilde{c}'\left(\boldsymbol{y}_{1:(t-1)}, \boldsymbol{a}_{t-1}\right)\right) \mathrm{d}\boldsymbol{a}_{t-1} \tag{4.37}$$

Taking the ratio of Equation (4.37) and (4.34) yields $\tilde{c}'\left(\boldsymbol{y}_{1:t}, \boldsymbol{\alpha}_t\right)$ in Equation (4.35). Moreover, for Equation (4.36)

$$\begin{aligned}\tilde{h}''\left(\boldsymbol{y}_{1:t}, \boldsymbol{\alpha}_t\right) = \exp\left(o\left(\boldsymbol{y}_{1:t-1}\right)\right) g_t\left(\boldsymbol{y}_t \mid \boldsymbol{\alpha}_t\right) \int f\left(\boldsymbol{\alpha}_t \mid \boldsymbol{a}_{t-1}\right) p\left(\boldsymbol{a}_{t-1} \mid \boldsymbol{y}_{1:t-1}\right) \\ \cdot \Bigg(\left(s_t'\left(\boldsymbol{\alpha}_t, \boldsymbol{a}_{t-1}\right) + \tilde{c}'\left(\boldsymbol{y}_{1:(t-1)}, \boldsymbol{a}_{t-1}\right)\right) \left(s_t'\left(\boldsymbol{\alpha}_t, \boldsymbol{a}_{t-1}\right) + \tilde{c}'\left(\boldsymbol{y}_{1:(t-1)}, \boldsymbol{a}_{t-1}\right)\right)^\top \\ + s_t''\left(\boldsymbol{\alpha}_t, \boldsymbol{a}_{t-1}\right) + \tilde{c}''\left(\boldsymbol{y}_{1:(t-1)}, \boldsymbol{a}_{t-1}\right) \Bigg) \mathrm{d}\boldsymbol{a}_{t-1}\end{aligned}$$

where we again take the ratio with (4.34). Unlike before, we need to evaluate two ratios, $\tilde{h}'\left(\boldsymbol{y}_{1:t}, \boldsymbol{a}_t\right) / \tilde{h}\left(\boldsymbol{y}_{1:t}, \boldsymbol{a}_t\right)$ and $\tilde{h}''\left(\boldsymbol{y}_{1:t}, \boldsymbol{a}_t\right) / \tilde{h}\left(\boldsymbol{y}_{1:t}, \boldsymbol{a}_t\right)$, which require evaluation of expressions of the form

$$\frac{\int f\left(\boldsymbol{\alpha}_t \mid \boldsymbol{a}_{t-1}\right) p\left(\boldsymbol{a}_{t-1} \mid \boldsymbol{y}_{1:t-1}\right) \kappa_t\left(\boldsymbol{\alpha}_t, \boldsymbol{a}_{t-1}\right) \mathrm{d}\boldsymbol{a}_{t-1}}{\int f\left(\boldsymbol{\alpha}_t \mid \boldsymbol{a}_{t-1}\right) p\left(\boldsymbol{a}_{t-1} \mid \boldsymbol{y}_{1:t-1}\right) \mathrm{d}\boldsymbol{a}_{t-1}} \tag{4.38}$$

for some function κ_t. To do so, redefine the weights in Equation (4.18) in the forward particle filter shown in Algorithm 2 to

$$w_t^{(i)} \propto \frac{g_t\left(\boldsymbol{y}_t \,\middle|\, \boldsymbol{\alpha}_t^{(i)}\right) \sum_{j=1}^N f\left(\boldsymbol{\alpha}_t^{(i)} \,\middle|\, \boldsymbol{\alpha}_{t-1}^{(j)}\right) w_{t-1}^{(j)}}{q\left(\boldsymbol{\alpha}_t^{(i)} \,\middle|\, \left\{\left(\boldsymbol{\alpha}_{t-1}^{(j)}, w_{t-1}^{(j)}\right)\right\}_{j=1,\dots,N}, \boldsymbol{y}_t\right)}$$

where we have made it explicit that the proposal distribution, q, may depend on the previous particle cloud and assume that we use the same number of particles at time 0. Further, we define the weights

$$\bar{w}_t^{(i,j)} = \frac{f\left(\boldsymbol{\alpha}_t^{(i)} \,\middle|\, \boldsymbol{\alpha}_{t-1}^{(j)}\right) w_{t-1}^{(j)}}{\sum_{k=1}^N f\left(\boldsymbol{\alpha}_t^{(i)} \,\middle|\, \boldsymbol{\alpha}_{t-1}^{(k)}\right) w_{t-1}^{(k)}} \tag{4.39}$$

Now a discrete approximation of the expression in Equation (4.38) for each particle i is given by

$$\sum_{j=1}^N \bar{w}_t^{(i,j)} \kappa_t\left(\boldsymbol{\alpha}_t^{(i)}, \boldsymbol{\alpha}_{t-1}^{(j)}\right)$$

Thus, the recursive formula for the gradient approximation is

$$\begin{aligned}\boldsymbol{\zeta}_t^{(i)} &= \sum_{j=1}^N \bar{w}_t^{(i,j)} \left(s_t'\left(\boldsymbol{\alpha}_t^{(i)}, \boldsymbol{\alpha}_{t-1}^{(j)}\right) + \boldsymbol{\zeta}_{t-1}^{(j)}\right) \\ \nabla o\left(\boldsymbol{y}_{1:t}\right) &\approx \sum_{i=1}^N w_t^{(i)} \boldsymbol{\zeta}_t^{(i)}\end{aligned} \tag{4.40}$$

and for the Hessian we have

$$\begin{aligned}\boldsymbol{\Upsilon}_t^{(i)} = \sum_{j=1}^N \bar{w}_t^{(i,j)} \Bigg(&\left(s_t'\left(\boldsymbol{\alpha}_t^{(i)}, \boldsymbol{\alpha}_{t-1}^{(j)}\right) + \boldsymbol{\zeta}_{t-1}^{(j)}\right)\left(s_t'\left(\boldsymbol{\alpha}_t^{(i)}, \boldsymbol{\alpha}_{t-1}^{(j)}\right) + \boldsymbol{\zeta}_{t-1}^{(j)}\right)^\top \\ &+ s_t''\left(\boldsymbol{\alpha}_t^{(i)}, \boldsymbol{\alpha}_{t-1}^{(j)}\right) + \boldsymbol{\Upsilon}_{t-1}^{(j)} \Bigg) - \boldsymbol{\zeta}_t^{(i)}\boldsymbol{\zeta}_t^{(i)\top}\end{aligned} \tag{4.41}$$

such that

$$\nabla^2 o\left(\boldsymbol{y}_{1:t}\right) \approx \sum_{i=1}^N w_t^{(i)}\left(\boldsymbol{\zeta}_t^{(i)}\boldsymbol{\zeta}_t^{(i)\top} + \boldsymbol{\Upsilon}_t^{(i)}\right) - \nabla o\left(\boldsymbol{y}_{1:t}\right)\nabla o\left(\boldsymbol{y}_{1:t}\right)^\top$$

The issue with the latter method is that the method has an $\mathcal{O}\left(N^2\right)$ computational complexity because of the sums in Equation (4.39), (4.40), and (4.41). Further, because there is no direct dependence between pairs of particles, an alternative type of particle filter can be used. A particular type of particle filters that are well suited for approximations like those in Equation (4.38) are the so-called independent particle filters suggested by Lin et al. (2005). The key point about these filters is that they use a proposal distribution that only depends on the observed outcome, $\boldsymbol{y}_t$, or also the previous particle cloud or a group of particles, but not any particular particle. Currently, the implementation supports the use by independence particle filters like the bootstrap filter using the mean of the previous particle cloud or a filter that makes a mode approximation using the mean of the previous particle cloud. Details are omitted for the sake of brevity, because the filters are very similar to those covered in Section 4.1.3 and 4.2.

An alternative to the methods in the **dynamichazard** package is the **mssm** package. It contains both the method shown in Cappé et al. (2005) and the method suggested by Poyiadjis et al. (2011), but for more general models. Moreover, the **mssm** package has an implementation of the dual k-d tree approximation method as in Klaas et al. (2006). This reduces the average-case complexity to $\mathcal{O}\left(N \log N\right)$, and thus it allows one to use substantially more particles. Lastly, the **mssm** package also allows for two types of antithetic variables like those suggested by Durbin and Koopman (1997). This decreases the variance of the estimates at a fixed computational cost.

Bibliography

Aalen, O. O. (1989). A linear regression model for the analysis of life times. *Statistics in Medicine*, 8(8):907–925.

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

Anderson, E., Bai, Z., Bischof, C., Blackford, S., Demmel, J., Dongarra, J., Du Croz, J., Greenbaum, A., Hammarling, S., McKenney, A., and Sorensen, D. (1999). *LAPACK users' guide*. Society for Industrial and Applied Mathematics, Philadelphia, PA, third edition.

Azizpour, S., Giesecke, K., and Schwenkler, G. (2018). Exploring the sources of default clustering. *Journal of Financial Economics*, 129(1):154 – 183.

BackBlaze (2014). Hard drive smart stats. Retrieved May 2017, https://www.backblaze.com/blog/hard-drive-smart-stats/.

BackBlaze (2017). Hard drive data and stats. Retrieved May 2017, https://www.backblaze.com/b2/hard-drive-test-data.html.

Bates, D. M., Mächler, M., Bolker, B., and Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1):1–48.

Beaver, W. H. (1966). Financial ratios as predictors of failure. *Journal of Accounting Research*, 4:71–111.

Berg, D. (2007). Bankruptcy prediction by generalized additive models. *Applied Stochastic Models in Business and Industry*, 23(2):129–143.

Berkowitz, J., Christoffersen, P., and Pelletier, D. (2011). Evaluating Value-at-Risk models with desk-level data. *Management Science*, 57(12):2213–2227.

Bharath, S. T. and Shumway, T. (2008). Forecasting default with the Merton distance to default model. *The Review of Financial Studies*, 21(3):1339–1369.

Board, O. A. R. (2013). *OpenMP application program interface*. Version 4.0.

Briers, M., Doucet, A., and Maskell, S. (2009). Smoothing algorithms for state–space models. *Annals of the Institute of Statistical Mathematics*, 62(1):61.

Broström, G. (2017). *eha: Event history analysis.* R package version 2.5.0.

Bühlmann, P. and Van De Geer, S. (2011). *Statistics for high-dimensional data: Methods, theory and applications.* Springer Science & Business Media.

Bühlmann, P. and Hothorn, T. (2007). Boosting algorithms: Regularization, prediction and model fitting. *Statistical Science*, 22(4):477–505.

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

Cappé, O., Moulines, E., and Ryden, T. (2005). *Inference in hidden Markov models.* Springer-Verlag New York.

Caruana, R. and Niculescu-Mizil, A. (2006). An empirical comparison of supervised learning algorithms. In *Proceedings of the 23rd International Conference on Machine Learning*, pages 161–168, New York, NY, USA. ACM.

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

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

Chen, P. and Wu, C. (2014). Default prediction with dynamic sectoral and macroeconomic frailties. *Journal of Banking & Finance*, 40:211–226.

Chen, T. and Guestrin, C. (2016). Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 785–794, New York, NY, USA. ACM.

Christoffersen, B. (2019). *dynamichazard: Dynamic hazard models using state space models.* R package version 0.6.4.

Christoffersen, B., Matin, R., and Mølgaard, P. (2019). Can machine learning models capture correlations in corporate distresses? Available at SSRN: `https://ssrn.com/abstract=3273985`.

Clements, M. and Liu, X.-R. (2016). *rstpm2: Generalized survival models.* R package version 1.3.4.

Cox, D. R. (1972). Regression models and life-tables. *Journal of the Royal Statistical Society Series B*, 34(2):187–220.

Danmarks Statistik (2018a). Fiks9: Firmaernes køb og salg, historisk sammendrag efter beløb. Retrieved September 2018, http://www.statistikbanken.dk/FIKS9.

Danmarks Statistik (2018b). Konk9: Erklærede konkurser (historisk sammendrag). Retrieved September 2018, http://www.statistikbanken.dk/KONK9.

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

Davidson-Pilon, C. (2019). lifelines: survival analysis in python. *Journal of Open Source Software*, 4(40).

Del Moral, P., Doucet, A., and Singh, S. (2010). Forward smoothing using sequential Monte Carlo. *arXiv preprint arXiv:1012.5390*.

Dempster, A. P., Laird, N. M., and Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society. Series B (Methodological)*, 39(1):1–38.

Douc, R. and Cappé, O. (2005). Comparison of resampling schemes for particle filtering. In *Image and Signal Processing and Analysis, 2005. ISPA 2005. Proceedings of the 4th International Symposium on*, pages 64–69. IEEE.

Doucet, A. and Johansen, A. M. (2009). A tutorial on particle filtering and smoothing: Fifteen years later. *Handbook of nonlinear filtering*, 12(656-704):3.

Duan, J.-C. and Fulop, A. (2013). Multiperiod corporate default prediction with the partially-conditioned forward intensity. Working paper.

Duan, J.-C., Sun, J., and Wang, T. (2012). Multiperiod corporate default prediction–A forward intensity approach. *Journal of Econometrics*, 170(1):191–209.

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

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

Durbin, J. and Koopman, S. J. (1997). Monte Carlo maximum likelihood estimation for non-Gaussian state space models. *Biometrika*, 84(3):669–684.

Durbin, J. and Koopman, S. J. (2000). Time series analysis of non-Gaussian observations based on state space models from both classical and bayesian perspectives. *Journal of the Royal Statistical Society Series B*, 62(1):3–56.

Durbin, J. and Koopman, S. J. (2012). *Time series analysis by state space methods*, volume 38. Oxford University Press, Oxford.

Fahrmeir, L. (1992). Posterior mode estimation by extended Kalman filtering for multivariate dynamic generalized linear models. *Journal of the American Statistical Association*, 87(418):501–509.

Fahrmeir, L. (1994). Dynamic modelling and penalized likelihood estimation for discrete time survival data. *Biometrika*, 81(2):317–330.

Fahrmeir, L. and Kaufmann, H. (1991). On Kalman filtering, posterior mode estimation and Fisher scoring in dynamic exponential family regression. *Metrika*, 38(1):37–60.

Fahrmeir, L. and Wagenpfeil, S. (1996). Smoothing hazard functions and time-varying effects in discrete duration and competing risks models. *Journal of the American Statistical Association*, 91(436):1584–1594.

Fearnhead, P., Wyncoll, D., and Tawn, J. (2010). A sequential smoothing algorithm with linear computational cost. *Biometrika*, 97(2):447–464.

Filipe, S. F., Grammatikos, T., and Michala, D. (2016). Forecasting distress in European SME portfolios. *Journal of Banking & Finance*, 64:112 – 135.

Fine, J. P., Yan, J., and Kosorok, M. R. (2004). Temporal process regression. *Biometrika*, 91:683–703.

Friedman, J., Hastie, T., and Tibshirani, R. (2010). Regularization paths for generalized linear models via coordinate descent. *Journal of Statistical Software*, 33(1):1–22.

Friedman, J. H. (2001). Greedy function approximation: A gradient boosting machine. *The Annals of Statistics*, 29(5):1189–1232.

Friedman, J. H. and Popescu, B. E. (2008). Predictive learning via rule ensembles. *Ann. Appl. Stat.*, 2(3):916–954.

Frumento, P. (2016). *pch: Piecewise constant hazards models for censored and truncated data.* R package version 1.3.

Goeman, J. J. (2010). L1 penalized estimation in the Cox proportional hazards model. *Biometrical Journal*, 52(1):70–84.

Hamilton, D. T. and Cantor, R. (2004). Rating transition and default rates conditioned on outlooks. *The Journal of Fixed Income*, 14(2):54–70.

Harrell Jr, F. E. (2017). *rms: Regression modeling strategies.* R package version 5.1-1.

Hartikainen, J., Solin, A., and Särkkä, S. (2011). *Optimal filtering with Kalman filters and smoothers. A manual for the MATLAB toolbox EKF/UKF.*

Harvey, A. C. and Phillips, G. D. A. (1979). Maximum likelihood estimation of regression models with autoregressive-moving average disturbances. *Biometrika*, 66(1):49.

Helske, J. (2017). KFAS: Exponential family state space models in R. *Journal of Statistical Software*, 78(10):1–39.

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

Holmes, E. E. (2013). Derivation of an EM algorithm for constrained and unconstrained multivariate autoregressive state-space (MARSS) models. *ArXiv e-prints*.

Jackson, C. (2016). flexsurv: A platform for parametric survival modeling in R. *Journal of Statistical Software*, 70(8):1–33.

Jensen, T., Lando, D., and Medhat, M. (2017). Cyclicality and firm size in private firm defaults. *International Journal of Central Banking*, 13(4):97–145.

Johansen, A. (2009). SMCTC: Sequential monte carlo in C++. *Journal of Statistical Software, Articles*, 30(6):1–41.

Jones, S., Johnstone, D., and Wilson, R. (2017). Predicting corporate bankruptcy: An evaluation of alternative statistical frameworks. *Journal of Business Finance & Accounting*, 44(1-2):3–34.

Julier, S. J. and Uhlmann, J. K. (1997). New extension of the Kalman filter to nonlinear systems. In *Signal Processing, Sensor Fusion, and Target Recognition VI*, volume 3068, Orlando, FL, United States.

Kantas, N., Doucet, A., Singh, S. S., Maciejowski, J., Chopin, N., et al. (2015). On particle methods for parameter estimation in state-space models. *Statistical Science*, 30(3):328–351.

Kim, M.-J. and Kang, D.-K. (2010). Ensemble with neural networks for bankruptcy prediction. *Expert Systems with Applications*, 37(4):3373–3379.

King, A., Nguyen, D., and Ionides, E. (2016). Statistical inference for partially observed Markov processes via the R package pomp. *Journal of Statistical Software*, 69(1):1–43.

King, A. A., Ionides, E. L., Bretó, C. M., Ellner, S. P., Ferrari, M. J., Kendall, B. E., Lavine, M., Nguyen, D., Reuman, D. C., Wearing, H., and Wood, S. N. (2017). *pomp: Statistical inference for partially observed Markov processes*. R package, version 1.15.

Kitagawa, G. (1994). The two-filter formula for smoothing and an implementation of the Gaussian-sum smoother. *Annals of the Institute of Statistical Mathematics*, 46(4):605–623.

Klaas, M., Briers, M., De Freitas, N., Doucet, A., Maskell, S., and Lang, D. (2006). Fast particle smoothing: If i had a million particles. In *Proceedings of the 23rd International Conference on Machine Learning*, pages 481–488, New York, NY, USA. ACM.

Klein, A. (2015). Csi: Backblaze – dissecting 3tb drive failure. Retrieved May 2017, https://www.backblaze.com/blog/3tb-hard-drive-failure/.

Klein, A. (2016). One billion drive hours and counting: Q1 2016 hard drive stats. Retrieved May 2017, https://www.backblaze.com/blog/hard-drive-reliability-stats-q1-2016/.

Kooperberg, C. (2015). *polspline: Polynomial spline routines*. R package version 1.1.12.

Koopman, S. J. and Durbin, J. (2000). Fast filtering and smoothing for multivariate state space models. *Journal of Time Series Analysis*, 21(3):281–296.

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

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

Koopman, S. J., Shephard, N., and Doornik, J. A. (2008). *Statistical algorithms for models in state space form: SsfPack 3.0.* Timberlake Consultants Press.

Kupiec, P. (1995). Techniques for verifying the accuracy of risk measurement models. *The Journal of Derivatives*, 3:73–84.

Kwon, T. Y. and Lee, Y. (2018). Industry specific defaults. *Journal of Empirical Finance*, 45:45–58.

Lambert, P. C. and Royston, P. (2009). Further development of flexible parametric models for survival analysis. *Stata Journal*, 9(2):265–290(26).

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

Lando, D. and Nielsen, M. S. (2010). Correlation in corporate defaults: Contagion or conditional independence? *Journal of Financial Intermediation*, 19(3):355–372.

Lin, M. T., Zhang, J. L., Cheng, Q., and Chen, R. (2005). Independent particle filters. *Journal of the American Statistical Association*, 100(472):1412–1421.

Liu, X.-R., Pawitan, Y., and Clements, M. (2016). Parametric and penalized generalized survival models. *Statistical Methods in Medical Research*, 27(5):1531–1546.

Liu, X.-R., Pawitan, Y., and Clements, M. S. (2017). Generalized survival models for correlated time-to-event data. *Statistics in Medicine*, 36(29):4743–4762.

Louis, T. A. (1982). Finding the observed information matrix when using the EM algorithm. *Journal of the Royal Statistical Society. Series B (Methodological)*, 44(2):226–233.

Malik, S. and Pitt, M. K. (2011). Particle filters for continuous likelihood evaluation and maximisation. *Journal of Econometrics*, 165(2):190–209.

Martinussen, T. and Scheike, T. H. (2006). *Dynamic regression models for survival data.* Springer-Verlag, NY.

Menegaz, H. M. T. (2016). *Unscented Kalman filtering on euclidean and riemannian manifolds.* PhD thesis, University of Brasília.

Meng, X.-L. and Rubin, D. B. (1993). Maximum likelihood estimation via the ECM algorithm: A general framework. *Biometrika*, 80(2):267–278.

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

Min, J. H. and Lee, Y.-C. (2005). Bankruptcy prediction using support vector machine with optimal choice of kernel function parameters. *Expert Systems with Applications*, 28(4):603–614.

Nordh, J. (2017). pyParticleEst: A Python framework for particle-based estimation methods. *Journal of Statistical Software, Articles*, 78(3):1–25.

Park, M. Y. and Hastie, T. (2013). *glmpath: L1 regularization path for generalized linear models and Cox proportional hazards model.* R package version 0.97.

Peng, J.-Y. and Aston, J. (2011). The state space models toolbox for MATLAB. *Journal of Statistical Software, Articles*, 41(6):1–26.

Petris, G. and Petrone, S. (2011). State space models in R. *Journal of Statistical Software*, 41(1):1–25.

Pinheiro, J. C. and Bates, D. M. (2000). *Mixed-effects models in S and S-PLUS.* Springer-Verlag New York.

Pitt, M. K. and Shephard, N. (1999). Filtering via simulation: Auxiliary particle filters. *Journal of the American Statistical Association*, 94(446):590–599.

Poyiadjis, G., Doucet, A., and Singh, S. S. (2011). Particle approximations of the score and observed information matrix in state space models with application to parameter estimation. *Biometrika*, 98(1):65–80.

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

R Core Team (2018). *R: A language and environment for statistical computing.* R Foundation for Statistical Computing, Vienna, Austria.

Rigatos, G. (2017). *State-space approaches for modelling and control in financial engineering, systems theory and machie learning methods.* Springer-Verlag, New York.

Rossum, G. (2017). *Python reference manual.* Release 2.7.14.

Sanderson, C. and Curtin, R. (2016). Armadillo: A template-based C++ library for linear algebra. *The Journal of Open Source Software*, 1(2).

SAS Institute Inc (2017). *SAS/STAT software, version 9.4.* Cary, NC.

Schön, T. B., Wills, A., and Ninness, B. (2011). System identification of nonlinear state-space models. *Automatica*, 47(1):39–49.

Schwaab, B., Koopman, S. J., and Lucas, A. (2017). Global credit risk: World, country and industry factors. *Journal of Applied Econometrics*, 32(2):296–317.

Shumway, R. H. and Stoffer, D. S. (1982). An approach to time series smoothing and forecasting using the EM algorithm. *Journal of Time Series Analysis*, 3(4):253–264.

Shumway, T. (2001). Forecasting bankruptcy more accurately: A simple hazard model. *The Journal of Business*, 74(1):101–124.

Simon, N., Friedman, J., Hastie, T., and Tibshirani, R. (2011). Regularization paths for Cox's proportional hazards model via coordinate descent. *Journal of Statistical Software*, 39(1):1–13.

StataCorp (2017). *Stata statistical software: Release 15.* StataCorp LLC, College Station, TX.

Team, S. D. (2017). *The Stan core library.* Cary, NC. Version 2.17.0.

Terry M. Therneau and Patricia M. Grambsch (2000). *Modeling survival data: Extending the Cox model.* Springer-Verlag, New York.

Therneau, T. M. (2015). *A package for survival analysis in S.*

Thomas, L. and Reyes, E. (2014). Tutorial: Survival estimation for Cox regression models with time-varying coefficients using SAS and R. *Journal of Statistical Software*, 61(1):1–23.

Tian, S., Yu, Y., and Guo, H. (2015). Variable selection and corporate bankruptcy forecasts. *Journal of Banking & Finance*, 52:89–100.

Tusell, F. (2011). Kalman filtering in R. *Journal of Statistical Software*, 39(1):1–27.

Tutz, G. and Schmid, M. (2016). *Modeling discrete time-to-event data.* Springer-Verlag, NY.

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

Vaupel, J. W., Manton, K. G., and Stallard, E. (1979). The impact of heterogeneity in individual frailty on the dynamics of mortality. *Demography*, 16(3):439–454.

Wan, E. A. and Merwe, R. V. D. (2000). The unscented Kalman filter for nonlinear estimation. In *Proceedings of the IEEE 2000 Adaptive Systems for Signal Processing, Communications, and Control Symposium (Cat. No.00EX373)*, pages 153–158.

Wang, X., Chen, M.-H., Wang, W., and Yan, J. (2017). *dynsurv: Dynamic models for survival data.* R package version 0.3-5.

Wood, S. (2017). *Generalized additive models: An introduction with R.* Chapman and Hall/CRC, 2 edition.

Wood, S. N. (2013). On p-values for smooth components of an extended generalized additive model. *Biometrika*, 100(1):221–228.

Wood, S. N., Goude, Y., and Shaw, S. (2015). Generalized additive models for large data sets. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 64(1):139–155.

Xianyi, Z., Qian, W., and Yunquan, Z. (2012). Model-driven level 3 BLAS performance optimization on loongson 3a processor. In *2012 IEEE 18th International Conference on Parallel and Distributed Systems*, pages 684–691.

Xiong, K., Zhang, H., and Chan, C. (2006). Performance evaluation of UKF-based nonlinear filtering. *Automatica*, 42(2):261–270.

Yan, J. and Fine, J. P. (2004). Estimating equations for association structures. *Statistics in Medicine*, 23:859–880.

Zhou, S. (2010). Thresholded Lasso for high dimensional variable selection and statistical estimation. Working paper.

Zhou, Y. (2015). vsmc: Parallel sequential Monte Carlo in C++. *Journal of Statistical Software, Articles*, 62(9):1–49.

Zięba, M., Tomczak, S. K., and Tomczak, J. M. (2016). Ensemble boosted trees with synthetic features generation in application to bankruptcy prediction. *Expert Systems with Applications*, 58:93–101.

TITLER I PH.D.SERIEN:

2004

1. Martin Grieger
   Internet-based Electronic Marketplaces and Supply Chain Management

2. Thomas Basbøll
   LIKENESS
   A Philosophical Investigation

3. Morten Knudsen
   Beslutningens vaklen
   En systemteoretisk analyse of moderniseringen af et amtskommunalt sundhedsvæsen 1980-2000

4. Lars Bo Jeppesen
   Organizing Consumer Innovation
   A product development strategy that is based on online communities and allows some firms to benefit from a distributed process of innovation by consumers

5. Barbara Dragsted
   SEGMENTATION IN TRANSLATION AND TRANSLATION MEMORY SYSTEMS
   An empirical investigation of cognitive segmentation and effects of integrating a TM system into the translation process

6. Jeanet Hardis
   Sociale partnerskaber
   Et socialkonstruktivistisk casestudie af partnerskabsaktørers virkelighedsopfattelse mellem identitet og legitimitet

7. Henriette Hallberg Thygesen
   System Dynamics in Action

8. Carsten Mejer Plath
   Strategisk Økonomistyring

9. Annemette Kjærgaard
   Knowledge Management as Internal Corporate Venturing
   – a Field Study of the Rise and Fall of a Bottom-Up Process

10. Knut Arne Hovdal
    De profesjonelle i endring
    Norsk ph.d., ej til salg gennem Samfundslitteratur

11. Søren Jeppesen
    Environmental Practices and Greening Strategies in Small Manufacturing Enterprises in South Africa
    – A Critical Realist Approach

12. Lars Frode Frederiksen
    Industriel forskningsledelse
    – på sporet af mønstre og samarbejde i danske forskningsintensive virksomheder

13. Martin Jes Iversen
    The Governance of GN Great Nordic
    – in an age of strategic and structural transitions 1939-1988

14. Lars Pynt Andersen
    The Rhetorical Strategies of Danish TV Advertising
    A study of the first fifteen years with special emphasis on genre and irony

15. Jakob Rasmussen
    Business Perspectives on E-learning

16. Sof Thrane
    The Social and Economic Dynamics of Networks
    – a Weberian Analysis of Three Formalised Horizontal Networks

17. Lene Nielsen
    Engaging Personas and Narrative Scenarios – a study on how a user-centered approach influenced the perception of the design process in the e-business group at AstraZeneca

18. S.J Valstad
    Organisationsidentitet
    Norsk ph.d., ej til salg gennem Samfundslitteratur

19. Thomas Lyse Hansen
Six Essays on Pricing and Weather risk in Energy Markets

20. Sabine Madsen
Emerging Methods – An Interpretive Study of ISD Methods in Practice

21. Evis Sinani
The Impact of Foreign Direct Investment on Efficiency, Productivity Growth and Trade: An Empirical Investigation

22. Bent Meier Sørensen
Making Events Work Or, How to Multiply Your Crisis

23. Pernille Schnoor
Brand Ethos
Om troværdige brand- og virksomhedsidentiteter i et retorisk og diskursteoretisk perspektiv

24. Sidsel Fabech
Von welchem Österreich ist hier die Rede?
Diskursive forhandlinger og magtkampe mellem rivaliserende nationale identitetskonstruktioner i østrigske pressediskurser

25. Klavs Odgaard Christensen
Sprogpolitik og identitetsdannelse i flersprogede forbundsstater
Et komparativt studie af Schweiz og Canada

26. Dana B. Minbaeva
Human Resource Practices and Knowledge Transfer in Multinational Corporations

27. Holger Højlund
Markedets politiske fornuft
Et studie af velfærdens organisering i perioden 1990-2003

28. Christine Mølgaard Frandsen
A.s erfaring
Om mellemværendets praktik i en transformation af mennesket og subjektiviteten

29. Sine Nørholm Just
The Constitution of Meaning
– A Meaningful Constitution?
Legitimacy, identity, and public opinion in the debate on the future of Europe

2005

1. Claus J. Varnes
Managing product innovation through rules – The role of formal and structured methods in product development

2. Helle Hedegaard Hein
Mellem konflikt og konsensus
– Dialogudvikling på hospitalsklinikker

3. Axel Rosenø
Customer Value Driven Product Innovation – A Study of Market Learning in New Product Development

4. Søren Buhl Pedersen
Making space
An outline of place branding

5. Camilla Funck Ellehave
Differences that Matter
An analysis of practices of gender and organizing in contemporary workplaces

6. Rigmor Madeleine Lond
Styring af kommunale forvaltninger

7. Mette Aagaard Andreassen
Supply Chain versus Supply Chain
Benchmarking as a Means to Managing Supply Chains

8. Caroline Aggestam-Pontoppidan
From an idea to a standard
The UN and the global governance of accountants' competence

9. Norsk ph.d.

10. Vivienne Heng Ker-ni
An Experimental Field Study on the

Effectiveness of Grocer Media Advertising
Measuring Ad Recall and Recognition, Purchase Intentions and Short-Term Sales

11. Allan Mortensen
Essays on the Pricing of Corporate Bonds and Credit Derivatives

12. Remo Stefano Chiari
Figure che fanno conoscere
Itinerario sull'idea del valore cognitivo e espressivo della metafora e di altri tropi da Aristotele e da Vico fino al cognitivismo contemporaneo

13. Anders McIlquham-Schmidt
Strategic Planning and Corporate Performance
An integrative research review and a meta-analysis of the strategic planning and corporate performance literature from 1956 to 2003

14. Jens Geersbro
The TDF – PMI Case
Making Sense of the Dynamics of Business Relationships and Networks

15 Mette Andersen
Corporate Social Responsibility in Global Supply Chains
Understanding the uniqueness of firm behaviour

16. Eva Boxenbaum
Institutional Genesis: Micro – Dynamic Foundations of Institutional Change

17. Peter Lund-Thomsen
Capacity Development, Environmental Justice NGOs, and Governance: The Case of South Africa

18. Signe Jarlov
Konstruktioner af offentlig ledelse

19. Lars Stæhr Jensen
Vocabulary Knowledge and Listening Comprehension in English as a Foreign Language
An empirical study employing data elicited from Danish EFL learners

20. Christian Nielsen
Essays on Business Reporting
Production and consumption of strategic information in the market for information

21. Marianne Thejls Fischer
Egos and Ethics of Management Consultants

22. Annie Bekke Kjær
Performance management i Proces-innovation
– belyst i et social-konstruktivistisk perspektiv

23. Suzanne Dee Pedersen
GENTAGELSENS METAMORFOSE
Om organisering af den kreative gøren i den kunstneriske arbejdspraksis

24. Benedikte Dorte Rosenbrink
Revenue Management
Økonomiske, konkurrencemæssige & organisatoriske konsekvenser

25. Thomas Riise Johansen
Written Accounts and Verbal Accounts
The Danish Case of Accounting and Accountability to Employees

26. Ann Fogelgren-Pedersen
The Mobile Internet: Pioneering Users' Adoption Decisions

27. Birgitte Rasmussen
Ledelse i fællesskab – de tillidsvalgtes fornyende rolle

28. Gitte Thit Nielsen
Remerger
– skabende ledelseskræfter i fusion og opkøb

29. Carmine Gioia
A MICROECONOMETRIC ANALYSIS OF MERGERS AND ACQUISITIONS

30. Ole Hinz
Den effektive forandringsleder: pilot, pædagog eller politiker?
Et studie i arbejdslederes meningstilskrivninger i forbindelse med vellykket gennemførelse af ledelsesinitierede forandringsprojekter

31. Kjell-Åge Gotvassli
Et praksisbasert perspektiv på dynamiske
læringsnettverk i toppidretten
Norsk ph.d., ej til salg gennem Samfundslitteratur

32. Henriette Langstrup Nielsen
Linking Healthcare
An inquiry into the changing performances of web-based technology for asthma monitoring

33. Karin Tweddell Levinsen
Virtuel Uddannelsespraksis
Master i IKT og Læring – et casestudie i hvordan proaktiv proceshåndtering kan forbedre praksis i virtuelle læringsmiljøer

34. Anika Liversage
Finding a Path
Labour Market Life Stories of Immigrant Professionals

35. Kasper Elmquist Jørgensen
Studier i samspillet mellem stat og erhvervsliv i Danmark under
1. verdenskrig

36. Finn Janning
A DIFFERENT STORY
Seduction, Conquest and Discovery

37. Patricia Ann Plackett
Strategic Management of the Radical Innovation Process
Leveraging Social Capital for Market Uncertainty Management

2006

1. Christian Vintergaard
Early Phases of Corporate Venturing

2. Niels Rom-Poulsen
Essays in Computational Finance

3. Tina Brandt Husman
Organisational Capabilities, Competitive Advantage & Project-Based Organisations
The Case of Advertising and Creative Good Production

4. Mette Rosenkrands Johansen
Practice at the top
– how top managers mobilise and use non-financial performance measures

5. Eva Parum
Corporate governance som strategisk kommunikations- og ledelsesværktøj

6. Susan Aagaard Petersen
Culture's Influence on Performance Management: The Case of a Danish Company in China

7. Thomas Nicolai Pedersen
The Discursive Constitution of Organizational Governance – Between unity and differentiation
The Case of the governance of environmental risks by World Bank environmental staff

8. Cynthia Selin
Volatile Visions: Transactons in Anticipatory Knowledge

9. Jesper Banghøj
Financial Accounting Information and Compensation in Danish Companies

10. Mikkel Lucas Overby
Strategic Alliances in Emerging High-Tech Markets: What's the Difference and does it Matter?

11. Tine Aage
External Information Acquisition of Industrial Districts and the Impact of Different Knowledge Creation Dimensions

A case study of the Fashion and Design Branch of the Industrial District of Montebelluna, NE Italy

12. Mikkel Flyverbom
Making the Global Information Society Governable
On the Governmentality of Multi-Stakeholder Networks

13. Anette Grønning
Personen bag
Tilstedevær i e-mail som interaktionsform mellem kunde og medarbejder i dansk forsikringskontekst

14. Jørn Helder
One Company – One Language?
The NN-case

15. Lars Bjerregaard Mikkelsen
Differing perceptions of customer value
Development and application of a tool for mapping perceptions of customer value at both ends of customer-supplier dyads in industrial markets

16. Lise Granerud
Exploring Learning
Technological learning within small manufacturers in South Africa

17. Esben Rahbek Pedersen
Between Hopes and Realities: Reflections on the Promises and Practices of Corporate Social Responsibility (CSR)

18. Ramona Samson
The Cultural Integration Model and European Transformation.
The Case of Romania

2007

1. Jakob Vestergaard
Discipline in The Global Economy
Panopticism and the Post-Washington Consensus

2. Heidi Lund Hansen
Spaces for learning and working
A qualitative study of change of work, management, vehicles of power and social practices in open offices

3. Sudhanshu Rai
Exploring the internal dynamics of software development teams during user analysis
A tension enabled Institutionalization Model; "Where process becomes the objective"

4. Norsk ph.d.
Ej til salg gennem Samfundslitteratur

5. Serden Ozcan
EXPLORING HETEROGENEITY IN ORGANIZATIONAL ACTIONS AND OUTCOMES
A Behavioural Perspective

6. Kim Sundtoft Hald
Inter-organizational Performance Measurement and Management in Action
– An Ethnography on the Construction of Management, Identity and Relationships

7. Tobias Lindeberg
Evaluative Technologies
Quality and the Multiplicity of Performance

8. Merete Wedell-Wedellsborg
Den globale soldat
Identitetsdannelse og identitetsledelse i multinationale militære organisationer

9. Lars Frederiksen
Open Innovation Business Models
Innovation in firm-hosted online user communities and inter-firm project ventures in the music industry
– A collection of essays

10. Jonas Gabrielsen
Retorisk toposlære – fra statisk 'sted' til persuasiv aktivitet

11. Christian Moldt-Jørgensen
Fra meningsløs til meningsfuld evaluering.
Anvendelsen af studentertilfredshedsmålinger på de korte og mellemlange videregående uddannelser set fra et psykodynamisk systemperspektiv

12. Ping Gao
Extending the application of actor-network theory
Cases of innovation in the telecommunications industry

13. Peter Mejlby
Frihed og fængsel, en del af den samme drøm?
Et phronetisk baseret casestudie af frigørelsens og kontrollens sameksistens i værdibaseret ledelse!

14. Kristina Birch
Statistical Modelling in Marketing

15. Signe Poulsen
Sense and sensibility:
The language of emotional appeals in insurance marketing

16. Anders Bjerre Trolle
Essays on derivatives pricing and dynamic asset allocation

17. Peter Feldhütter
Empirical Studies of Bond and Credit Markets

18. Jens Henrik Eggert Christensen
Default and Recovery Risk Modeling and Estimation

19. Maria Theresa Larsen
Academic Enterprise: A New Mission for Universities or a Contradiction in Terms?
Four papers on the long-term implications of increasing industry involvement and commercialization in academia

20. Morten Wellendorf
Postimplementering af teknologi i den offentlige forvaltning
Analyser af en organisations kontinuerlige arbejde med informationsteknologi

21. Ekaterina Mhaanna
Concept Relations for Terminological Process Analysis

22. Stefan Ring Thorbjørnsen
Forsvaret i forandring
Et studie i officerers kapabiliteter under påvirkning af omverdenens forandringspres mod øget styring og læring

23. Christa Breum Amhøj
Det selvskabte medlemskab om managementstaten, dens styringsteknologier og indbyggere

24. Karoline Bromose
Between Technological Turbulence and Operational Stability
– An empirical case study of corporate venturing in TDC

25. Susanne Justesen
Navigating the Paradoxes of Diversity in Innovation Practice
– A Longitudinal study of six very different innovation processes – in practice

26. Luise Noring Henler
Conceptualising successful supply chain partnerships
– Viewing supply chain partnerships from an organisational culture perspective

27. Mark Mau
Kampen om telefonen
Det danske telefonvæsen under den tyske besættelse 1940-45

28. Jakob Halskov
The semiautomatic expansion of existing terminological ontologies using knowledge patterns discovered

on the WWW – an implementation and evaluation

29. Gergana Koleva
European Policy Instruments Beyond Networks and Structure: The Innovative Medicines Initiative

30. Christian Geisler Asmussen
Global Strategy and International Diversity: A Double-Edged Sword?

31. Christina Holm-Petersen
Stolthed og fordom
Kultur- og identitetsarbejde ved skabelsen af en ny sengeafdeling gennem fusion

32. Hans Peter Olsen
Hybrid Governance of Standardized States
Causes and Contours of the Global Regulation of Government Auditing

33. Lars Bøge Sørensen
Risk Management in the Supply Chain

34. Peter Aagaard
Det unikkes dynamikker
De institutionelle mulighedsbetingelser bag den individuelle udforskning i professionelt og frivilligt arbejde

35. Yun Mi Antorini
Brand Community Innovation
An Intrinsic Case Study of the Adult Fans of LEGO Community

36. Joachim Lynggaard Boll
Labor Related Corporate Social Performance in Denmark
Organizational and Institutional Perspectives

2008

1. Frederik Christian Vinten
Essays on Private Equity

2. Jesper Clement
Visual Influence of Packaging Design on In-Store Buying Decisions

3. Marius Brostrøm Kousgaard
Tid til kvalitetsmåling?
– Studier af indrulleringsprocesser i forbindelse med introduktionen af kliniske kvalitetsdatabaser i speciallægepraksissektoren

4. Irene Skovgaard Smith
Management Consulting in Action Value creation and ambiguity in client-consultant relations

5. Anders Rom
Management accounting and integrated information systems
How to exploit the potential for management accounting of information technology

6. Marina Candi
Aesthetic Design as an Element of Service Innovation in New Technology-based Firms

7. Morten Schnack
Teknologi og tværfaglighed
– en analyse af diskussionen omkring indførelse af EPJ på en hospitalsafdeling

8. Helene Balslev Clausen
Juntos pero no revueltos – un estudio sobre emigrantes norteamericanos en un pueblo mexicano

9. Lise Justesen
Kunsten at skrive revisionsrapporter.
En beretning om forvaltningsrevisionens beretninger

10. Michael E. Hansen
The politics of corporate responsibility: CSR and the governance of child labor and core labor rights in the 1990s

11. Anne Roepstorff
Holdning for handling – en etnologisk undersøgelse af Virksomheders Sociale Ansvar/CSR

12. Claus Bajlum
Essays on Credit Risk and Credit Derivatives

13. Anders Bojesen
The Performative Power of Competence – an Inquiry into Subjectivity and Social Technologies at Work

14. Satu Reijonen
Green and Fragile
A Study on Markets and the Natural Environment

15. Ilduara Busta
Corporate Governance in Banking
A European Study

16. Kristian Anders Hvass
A Boolean Analysis Predicting Industry Change: Innovation, Imitation & Business Models
The Winning Hybrid: A case study of isomorphism in the airline industry

17. Trine Paludan
De uvidende og de udviklingsparate
Identitet som mulighed og restriktion blandt fabriksarbejdere på det aftayloriserede fabriksgulv

18. Kristian Jakobsen
Foreign market entry in transition economies: Entry timing and mode choice

19. Jakob Elming
Syntactic reordering in statistical machine translation

20. Lars Brømsøe Termansen
Regional Computable General Equilibrium Models for Denmark
Three papers laying the foundation for regional CGE models with agglomeration characteristics

21. Mia Reinholt
The Motivational Foundations of Knowledge Sharing

22. Frederikke Krogh-Meibom
The Co-Evolution of Institutions and Technology
– A Neo-Institutional Understanding of Change Processes within the Business Press – the Case Study of Financial Times

23. Peter D. Ørberg Jensen
OFFSHORING OF ADVANCED AND HIGH-VALUE TECHNICAL SERVICES: ANTECEDENTS, PROCESS DYNAMICS AND FIRMLEVEL IMPACTS

24. Pham Thi Song Hanh
Functional Upgrading, Relational Capability and Export Performance of Vietnamese Wood Furniture Producers

25. Mads Vangkilde
Why wait?
An Exploration of first-mover advantages among Danish e-grocers through a resource perspective

26. Hubert Buch-Hansen
Rethinking the History of European Level Merger Control
A Critical Political Economy Perspective

2009

1. Vivian Lindhardsen
From Independent Ratings to Communal Ratings: A Study of CWA Raters' Decision-Making Behaviours

2. Guðrið Weihe
Public-Private Partnerships: Meaning and Practice

3. Chris Nøkkentved
Enabling Supply Networks with Collaborative Information Infrastructures
An Empirical Investigation of Business Model Innovation in Supplier Relationship Management

4. Sara Louise Muhr
Wound, Interrupted – On the Vulnerability of Diversity Management

5. Christine Sestoft
Forbrugeradfærd i et Stats- og Livsformsteoretisk perspektiv

6. Michael Pedersen
Tune in, Breakdown, and Reboot: On the production of the stress-fit self-managing employee

7. Salla Lutz
Position and Reposition in Networks
– Exemplified by the Transformation of the Danish Pine Furniture Manufacturers

8. Jens Forssbæck
Essays on market discipline in commercial and central banking

9. Tine Murphy
Sense from Silence – A Basis for Organised Action
How do Sensemaking Processes with Minimal Sharing Relate to the Reproduction of Organised Action?

10. Sara Malou Strandvad
Inspirations for a new sociology of art: A sociomaterial study of development processes in the Danish film industry

11. Nicolaas Mouton
On the evolution of social scientific metaphors:
A cognitive-historical enquiry into the divergent trajectories of the idea that collective entities – states and societies, cities and corporations – are biological organisms.

12. Lars Andreas Knutsen
Mobile Data Services:
Shaping of user engagements

13. Nikolaos Theodoros Korfiatis
Information Exchange and Behavior
A Multi-method Inquiry on Online Communities

14. Jens Albæk
Forestillinger om kvalitet og tværfaglighed på sygehuse
– skabelse af forestillinger i læge- og plejegrupperne angående relevans af nye idéer om kvalitetsudvikling gennem tolkningsprocesser

15. Maja Lotz
The Business of Co-Creation – and the Co-Creation of Business

16. Gitte P. Jakobsen
Narrative Construction of Leader Identity in a Leader Development Program Context

17. Dorte Hermansen
"Living the brand" som en brandorienteret dialogisk praxis:
Om udvikling af medarbejdernes brandorienterede dømmekraft

18. Aseem Kinra
Supply Chain (logistics) Environmental Complexity

19. Michael Nørager
How to manage SMEs through the transformation from non innovative to innovative?

20. Kristin Wallevik
Corporate Governance in Family Firms
The Norwegian Maritime Sector

21. Bo Hansen Hansen
Beyond the Process
Enriching Software Process Improvement with Knowledge Management

22. Annemette Skot-Hansen
Franske adjektivisk afledte adverbier, der tager præpositionssyntagmer indledt med præpositionen à som argumenter
En valensgrammatisk undersøgelse

23. Line Gry Knudsen
Collaborative R&D Capabilities
In Search of Micro-Foundations

24. Christian Scheuer
Employers meet employees
Essays on sorting and globalization

25. Rasmus Johnsen
The Great Health of Melancholy
A Study of the Pathologies of Performativity

26. Ha Thi Van Pham
Internationalization, Competitiveness Enhancement and Export Performance of Emerging Market Firms:
Evidence from Vietnam

27. Henriette Balieu
Kontrolbegrebets betydning for kausativalternationen i spansk
En kognitiv-typologisk analyse

2010

1. Yen Tran
Organizing Innovationin Turbulent Fashion Market
Four papers on how fashion firms create and appropriate innovation value

2. Anders Raastrup Kristensen
Metaphysical Labour
Flexibility, Performance and Commitment in Work-Life Management

3. Margrét Sigrún Sigurdardottir
Dependently independent
Co-existence of institutional logics in the recorded music industry

4. Ásta Dis Óladóttir
Internationalization from a small domestic base:
An empirical analysis of Economics and Management

5. Christine Secher
E-deltagelse i praksis – politikernes og forvaltningens medkonstruktion og konsekvenserne heraf

6. Marianne Stang Våland
What we talk about when we talk about space:
End User Participation between Processes of Organizational and Architectural Design

7. Rex Degnegaard
Strategic Change Management
Change Management Challenges in the Danish Police Reform

8. Ulrik Schultz Brix
Værdi i rekruttering – den sikre beslutning
En pragmatisk analyse af perception og synliggørelse af værdi i rekrutterings- og udvælgelsesarbejdet

9. Jan Ole Similä
Kontraktsledelse
Relasjonen mellom virksomhetsledelse og kontraktshåndtering, belyst via fire norske virksomheter

10. Susanne Boch Waldorff
Emerging Organizations: In between local translation, institutional logics and discourse

11. Brian Kane
Performance Talk
Next Generation Management of Organizational Performance

12. Lars Ohnemus
Brand Thrust: Strategic Branding and Shareholder Value
An Empirical Reconciliation of two Critical Concepts

13. Jesper Schlamovitz
Håndtering af usikkerhed i film- og byggeprojekter

14. Tommy Moesby-Jensen
Det faktiske livs forbindtlighed
Førsokratisk informeret, ny-aristotelisk ἦθος-tænkning hos Martin Heidegger

15. Christian Fich
Two Nations Divided by Common Values
French National Habitus and the Rejection of American Power

16. Peter Beyer
Processer, sammenhængskraft og fleksibilitet
Et empirisk casestudie af omstillingsforløb i fire virksomheder

17. Adam Buchhorn
Markets of Good Intentions
Constructing and Organizing Biogas Markets Amid Fragility and Controversy

18. Cecilie K. Moesby-Jensen
Social læring og fælles praksis
Et mixed method studie, der belyser læringskonsekvenser af et lederkursus for et praksisfællesskab af offentlige mellemledere

19. Heidi Boye
Fødevarer og sundhed i senmodernismen
– En indsigt i hyggefænomenet og de relaterede fødevarepraksisser

20. Kristine Munkgård Pedersen
Flygtige forbindelser og midlertidige mobiliseringer
Om kulturel produktion på Roskilde Festival

21. Oliver Jacob Weber
Causes of Intercompany Harmony in Business Markets – An Empirical Investigation from a Dyad Perspective

22. Susanne Ekman
Authority and Autonomy
Paradoxes of Modern Knowledge Work

23. Anette Frey Larsen
Kvalitetsledelse på danske hospitaler
– Ledelsernes indflydelse på introduktion og vedligeholdelse af kvalitetsstrategier i det danske sundhedsvæsen

24. Toyoko Sato
Performativity and Discourse: Japanese Advertisements on the Aesthetic Education of Desire

25. Kenneth Brinch Jensen
Identifying the Last Planner System
Lean management in the construction industry

26. Javier Busquets
Orchestrating Network Behavior for Innovation

27. Luke Patey
The Power of Resistance: India's National Oil Company and International Activism in Sudan

28. Mette Vedel
Value Creation in Triadic Business Relationships. Interaction, Interconnection and Position

29. Kristian Tørning
Knowledge Management Systems in Practice – A Work Place Study

30. Qingxin Shi
An Empirical Study of Thinking Aloud Usability Testing from a Cultural Perspective

31. Tanja Juul Christiansen
Corporate blogging: Medarbejderes kommunikative handlekraft

32. Malgorzata Ciesielska
Hybrid Organisations.
A study of the Open Source – business setting

33. Jens Dick-Nielsen
Three Essays on Corporate Bond Market Liquidity

34. Sabrina Speiermann
Modstandens Politik
Kampagnestyring i Velfærdsstaten.
En diskussion af trafikkampagners styringspotentiale

35. Julie Uldam
Fickle Commitment. Fostering political engagement in 'the flighty world of online activism'

36. Annegrete Juul Nielsen
Traveling technologies and transformations in health care

37. Athur Mühlen-Schulte
Organising Development
Power and Organisational Reform in the United Nations Development Programme

38. Louise Rygaard Jonas
Branding på butiksgulvet
Et case-studie af kultur- og identitetsarbejdet i Kvickly

2011

1. Stefan Fraenkel
Key Success Factors for Sales Force Readiness during New Product Launch
A Study of Product Launches in the Swedish Pharmaceutical Industry

2. Christian Plesner Rossing
International Transfer Pricing in Theory and Practice

3. Tobias Dam Hede
Samtalekunst og ledelsesdisciplin
– en analyse af coachingsdiskursens genealogi og governmentality

4. Kim Pettersson
Essays on Audit Quality, Auditor Choice, and Equity Valuation

5. Henrik Merkelsen
The expert-lay controversy in risk research and management. Effects of institutional distances. Studies of risk definitions, perceptions, management and communication

6. Simon S. Torp
Employee Stock Ownership:
Effect on Strategic Management and Performance

7. Mie Harder
Internal Antecedents of Management Innovation

8. Ole Helby Petersen
Public-Private Partnerships: Policy and Regulation – With Comparative and Multi-level Case Studies from Denmark and Ireland

9. Morten Krogh Petersen
'Good' Outcomes. Handling Multiplicity in Government Communication

10. Kristian Tangsgaard Hvelplund
Allocation of cognitive resources in translation - an eye-tracking and key-logging study

11. Moshe Yonatany
The Internationalization Process of Digital Service Providers

12. Anne Vestergaard
Distance and Suffering
Humanitarian Discourse in the age of Mediatization

13. Thorsten Mikkelsen
Personligsheds indflydelse på forretningsrelationer

14. Jane Thostrup Jagd
Hvorfor fortsætter fusionsbølgen udover "the tipping point"?
– en empirisk analyse af information og kognitioner om fusioner

15. Gregory Gimpel
Value-driven Adoption and Consumption of Technology: Understanding Technology Decision Making

16. Thomas Stengade Sønderskov
Den nye mulighed
Social innovation i en forretningsmæssig kontekst

17. Jeppe Christoffersen
Donor supported strategic alliances in developing countries

18. Vibeke Vad Baunsgaard
Dominant Ideological Modes of Rationality: Cross functional

integration in the process of product innovation

19. Throstur Olaf Sigurjonsson
Governance Failure and Icelands's Financial Collapse

20. Allan Sall Tang Andersen
Essays on the modeling of risks in interest-rate and inflation markets

21. Heidi Tscherning
Mobile Devices in Social Contexts

22. Birgitte Gorm Hansen
Adapting in the Knowledge Economy
Lateral Strategies for Scientists and Those Who Study Them

23. Kristina Vaarst Andersen
Optimal Levels of Embeddedness
The Contingent Value of Networked Collaboration

24. Justine Grønbæk Pors
Noisy Management
A History of Danish School Governing from 1970-2010

25. Stefan Linder
Micro-foundations of Strategic Entrepreneurship
Essays on Autonomous Strategic Action

26. Xin Li
Toward an Integrative Framework of National Competitiveness
An application to China

27. Rune Thorbjørn Clausen
Værdifuld arkitektur
Et eksplorativt studie af bygningers rolle i virksomheders værdiskabelse

28. Monica Viken
Markedsundersøkelser som bevis i varemerke- og markedsføringsrett

29. Christian Wymann
Tattooing
The Economic and Artistic Constitution of a Social Phenomenon

30. Sanne Frandsen
Productive Incoherence
A Case Study of Branding and Identity Struggles in a Low-Prestige Organization

31. Mads Stenbo Nielsen
Essays on Correlation Modelling

32. Ivan Häuser
Følelse og sprog
Etablering af en ekspressiv kategori, eksemplificeret på russisk

33. Sebastian Schwenen
Security of Supply in Electricity Markets

2012

1. Peter Holm Andreasen
The Dynamics of Procurement Management
- A Complexity Approach

2. Martin Haulrich
Data-Driven Bitext Dependency Parsing and Alignment

3. Line Kirkegaard
Konsulenten i den anden nat
En undersøgelse af det intense arbejdsliv

4. Tonny Stenheim
Decision usefulness of goodwill under IFRS

5. Morten Lind Larsen
Produktivitet, vækst og velfærd
Industrirådet og efterkrigstidens Danmark 1945 - 1958

6. Petter Berg
Cartel Damages and Cost Asymmetries

7. Lynn Kahle
Experiential Discourse in Marketing
A methodical inquiry into practice and theory

8. Anne Roelsgaard Obling
Management of Emotions in Accelerated Medical Relationships

9. Thomas Frandsen
Managing Modularity of
Service Processes Architecture

10. Carina Christine Skovmøller
CSR som noget særligt
Et casestudie om styring og meningsskabelse
i relation til CSR ud fra en
intern optik

11. Michael Tell
Fradragsbeskæring af selskabers
finansieringsudgifter
En skatteretlig analyse af SEL §§ 11,
11B og 11C

12. Morten Holm
Customer Profitability Measurement
Models
Their Merits and Sophistication
across Contexts

13. Katja Joo Dyppel
Beskatning af derivater
En analyse af dansk skatteret

14. Esben Anton Schultz
Essays in Labor Economics
Evidence from Danish Micro Data

15. Carina Risvig Hansen
"Contracts not covered, or not fully
covered, by the Public Sector Directive"

16. Anja Svejgaard Pors
Iværksættelse af kommunikation
- patientfigurer i hospitalets strategiske
kommunikation

17. Frans Bévort
Making sense of management with
logics
An ethnographic study of accountants
who become managers

18. René Kallestrup
The Dynamics of Bank and Sovereign
Credit Risk

19. Brett Crawford
Revisiting the Phenomenon of Interests
in Organizational Institutionalism
The Case of U.S. Chambers of
Commerce

20. Mario Daniele Amore
Essays on Empirical Corporate Finance

21. Arne Stjernholm Madsen
The evolution of innovation strategy
Studied in the context of medical
device activities at the pharmaceutical
company Novo Nordisk A/S in the
period 1980-2008

22. Jacob Holm Hansen
Is Social Integration Necessary for
Corporate Branding?
A study of corporate branding
strategies at Novo Nordisk

23. Stuart Webber
Corporate Profit Shifting and the
Multinational Enterprise

24. Helene Ratner
Promises of Reflexivity
Managing and Researching
Inclusive Schools

25. Therese Strand
The Owners and the Power: Insights
from Annual General Meetings

26. Robert Gavin Strand
In Praise of Corporate Social
Responsibility Bureaucracy

27. Nina Sormunen
Auditor's going-concern reporting
Reporting decision and content of the
report

28. John Bang Mathiasen
Learning within a product development
working practice:
- an understanding anchored
in pragmatism

29. Philip Holst Riis
Understanding Role-Oriented Enterprise
Systems: From Vendors to Customers

30. Marie Lisa Dacanay
Social Enterprises and the Poor
Enhancing Social Entrepreneurship and
Stakeholder Theory

31. Fumiko Kano Glückstad
Bridging Remote Cultures: Cross-lingual concept mapping based on the information receiver's prior-knowledge

32. Henrik Barslund Fosse
Empirical Essays in International Trade

33. Peter Alexander Albrecht
Foundational hybridity and its reproduction
Security sector reform in Sierra Leone

34. Maja Rosenstock
CSR - hvor svært kan det være?
Kulturanalytisk casestudie om udfordringer og dilemmaer med at forankre Coops CSR-strategi

35. Jeanette Rasmussen
Tweens, medier og forbrug
Et studie af 10-12 årige danske børns brug af internettet, opfattelse og forståelse af markedsføring og forbrug

36. Ib Tunby Gulbrandsen
'This page is not intended for a US Audience'
A five-act spectacle on online communication, collaboration & organization.

37. Kasper Aalling Teilmann
Interactive Approaches to Rural Development

38. Mette Mogensen
The Organization(s) of Well-being and Productivity
(Re)assembling work in the Danish Post

39. Søren Friis Møller
From Disinterestedness to Engagement
Towards Relational Leadership In the Cultural Sector

40. Nico Peter Berhausen
Management Control, Innovation and Strategic Objectives – Interactions and Convergence in Product Development Networks

41. Balder Onarheim
Creativity under Constraints
Creativity as Balancing 'Constrainedness'

42. Haoyong Zhou
Essays on Family Firms

43. Elisabeth Naima Mikkelsen
Making sense of organisational conflict
An empirical study of enacted sense-making in everyday conflict at work

2013

1. Jacob Lyngsie
Entrepreneurship in an Organizational Context

2. Signe Groth-Brodersen
Fra ledelse til selvet
En socialpsykologisk analyse af forholdet imellem selvledelse, ledelse og stress i det moderne arbejdsliv

3. Nis Høyrup Christensen
Shaping Markets: A Neoinstitutional Analysis of the Emerging Organizational Field of Renewable Energy in China

4. Christian Edelvold Berg
As a matter of size
THE IMPORTANCE OF CRITICAL MASS AND THE CONSEQUENCES OF SCARCITY FOR TELEVISION MARKETS

5. Christine D. Isakson
Coworker Influence and Labor Mobility
Essays on Turnover, Entrepreneurship and Location Choice in the Danish Maritime Industry

6. Niels Joseph Jerne Lennon
Accounting Qualities in Practice
Rhizomatic stories of representational faithfulness, decision making and control

7. Shannon O'Donnell
Making Ensemble Possible
How special groups organize for collaborative creativity in conditions of spatial variability and distance

8. Robert W. D. Veitch
Access Decisions in a
Partly-Digital World
Comparing Digital Piracy and Legal
Modes for Film and Music

9. Marie Mathiesen
Making Strategy Work
An Organizational Ethnography

10. Arisa Shollo
The role of business intelligence in
organizational decision-making

11. Mia Kaspersen
The construction of social and
environmental reporting

12. Marcus Møller Larsen
The organizational design of offshoring

13. Mette Ohm Rørdam
EU Law on Food Naming
The prohibition against misleading
names in an internal market context

14. Hans Peter Rasmussen
GIV EN GED!
Kan giver-idealtyper forklare støtte
til velgørenhed og understøtte
relationsopbygning?

15. Ruben Schachtenhaufen
Fonetisk reduktion i dansk

16. Peter Koerver Schmidt
Dansk CFC-beskatning
I et internationalt og komparativt
perspektiv

17. Morten Froholdt
Strategi i den offentlige sektor
En kortlægning af styringsmæssig
kontekst, strategisk tilgang, samt
anvendte redskaber og teknologier for
udvalgte danske statslige styrelser

18. Annette Camilla Sjørup
Cognitive effort in metaphor translation
An eye-tracking and key-logging study

19. Tamara Stucchi
The Internationalization
of Emerging Market Firms:
A Context-Specific Study

20. Thomas Lopdrup-Hjorth
"Let's Go Outside":
The Value of Co-Creation

21. Ana Alačovska
Genre and Autonomy in Cultural
Production
The case of travel guidebook
production

22. Marius Gudmand-Høyer
Stemningssindssygdommenes historie
i det 19. århundrede
Omtydningen af melankolien og
manien som bipolære stemningslidelser
i dansk sammenhæng under hensyn til
dannelsen af det moderne følelseslivs
relative autonomi.
En problematiserings- og erfarings-
analytisk undersøgelse

23. Lichen Alex Yu
Fabricating an S&OP Process
Circulating References and Matters
of Concern

24. Esben Alfort
The Expression of a Need
Understanding search

25. Trine Pallesen
Assembling Markets for Wind Power
An Inquiry into the Making of
Market Devices

26. Anders Koed Madsen
Web-Visions
Repurposing digital traces to organize
social attention

27. Lærke Højgaard Christiansen
BREWING ORGANIZATIONAL
RESPONSES TO INSTITUTIONAL LOGICS

28. Tommy Kjær Lassen
EGENTLIG SELVLEDELSE
En ledelsesfilosofisk afhandling om
selvledelsens paradoksale dynamik og
eksistentielle engagement

29. Morten Rossing
Local Adaption and Meaning Creation in Performance Appraisal

30. Søren Obed Madsen
Lederen som oversætter
Et oversættelsesteoretisk perspektiv på strategisk arbejde

31. Thomas Høgenhaven
Open Government Communities
Does Design Affect Participation?

32. Kirstine Zinck Pedersen
Failsafe Organizing?
A Pragmatic Stance on Patient Safety

33. Anne Petersen
Hverdagslogikker i psykiatrisk arbejde
En institutionsetnografisk undersøgelse af hverdagen i psykiatriske organisationer

34. Didde Maria Humle
Fortællinger om arbejde

35. Mark Holst-Mikkelsen
Strategieksekvering i praksis – barrierer og muligheder!

36. Malek Maalouf
Sustaining lean
Strategies for dealing with organizational paradoxes

37. Nicolaj Tofte Brenneche
Systemic Innovation In The Making
The Social Productivity of Cartographic Crisis and Transitions in the Case of SEEIT

38. Morten Gylling
The Structure of Discourse
A Corpus-Based Cross-Linguistic Study

39. Binzhang YANG
Urban Green Spaces for Quality Life - Case Study: the landscape architecture for people in Copenhagen

40. Michael Friis Pedersen
Finance and Organization:
The Implications for Whole Farm Risk Management

41. Even Fallan
Issues on supply and demand for environmental accounting information

42. Ather Nawaz
Website user experience
A cross-cultural study of the relation between users´ cognitive style, context of use, and information architecture of local websites

43. Karin Beukel
The Determinants for Creating Valuable Inventions

44. Arjan Markus
External Knowledge Sourcing and Firm Innovation
Essays on the Micro-Foundations of Firms’ Search for Innovation

2014

1. Solon Moreira
Four Essays on Technology Licensing and Firm Innovation

2. Karin Strzeletz Ivertsen
Partnership Drift in Innovation Processes
A study of the Think City electric car development

3. Kathrine Hoffmann Pii
Responsibility Flows in Patient-centred Prevention

4. Jane Bjørn Vedel
Managing Strategic Research
An empirical analysis of science-industry collaboration in a pharmaceutical company

5. Martin Gylling
Processuel strategi i organisationer
Monografi om dobbeltheden i tænkning af strategi, dels som vidensfelt i organisationsteori, dels som kunstnerisk tilgang til at skabe i erhvervsmæssig innovation

6. Linne Marie Lauesen
Corporate Social Responsibility in the Water Sector: How Material Practices and their Symbolic and Physical Meanings Form a Colonising Logic

7. Maggie Qiuzhu Mei
LEARNING TO INNOVATE: The role of ambidexterity, standard, and decision process

8. Inger Høedt-Rasmussen
Developing Identity for Lawyers Towards Sustainable Lawyering

9. Sebastian Fux
Essays on Return Predictability and Term Structure Modelling

10. Thorbjørn N. M. Lund-Poulsen
Essays on Value Based Management

11. Oana Brindusa Albu
Transparency in Organizing: A Performative Approach

12. Lena Olaison
Entrepreneurship at the limits

13. Hanne Sørum
DRESSED FOR WEB SUCCESS? An Empirical Study of Website Quality in the Public Sector

14. Lasse Folke Henriksen
Knowing networks How experts shape transnational governance

15. Maria Halbinger
Entrepreneurial Individuals Empirical Investigations into Entrepreneurial Activities of Hackers and Makers

16. Robert Spliid
Kapitalfondenes metoder og kompetencer

17. Christiane Stelling
Public-private partnerships & the need, development and management of trusting A processual and embedded exploration

18. Marta Gasparin
Management of design as a translation process

19. Kåre Moberg
Assessing the Impact of Entrepreneurship Education From ABC to PhD

20. Alexander Cole
Distant neighbors Collective learning beyond the cluster

21. Martin Møller Boje Rasmussen
Is Competitiveness a Question of Being Alike? How the United Kingdom, Germany and Denmark Came to Compete through their Knowledge Regimes from 1993 to 2007

22. Anders Ravn Sørensen
Studies in central bank legitimacy, currency and national identity Four cases from Danish monetary history

23. Nina Bellak
Can Language be Managed in International Business? Insights into Language Choice from a Case Study of Danish and Austrian Multinational Corporations (MNCs)

24. Rikke Kristine Nielsen
Global Mindset as Managerial Meta-competence and Organizational Capability: Boundary-crossing Leadership Cooperation in the MNC The Case of 'Group Mindset' in Solar A/S.

25. Rasmus Koss Hartmann
User Innovation inside government Towards a critically performative foundation for inquiry

26. Kristian Gylling Olesen
Flertydig og emergerende ledelse i folkeskolen
Et aktør-netværksteoretisk ledelsesstudie af politiske evalueringsreformers betydning for ledelse i den danske folkeskole

27. Troels Riis Larsen
Kampen om Danmarks omdømme 1945-2010
Omdømmearbejde og omdømmepolitik

28. Klaus Majgaard
Jagten på autenticitet i offentlig styring

29. Ming Hua Li
Institutional Transition and Organizational Diversity:
Differentiated internationalization strategies of emerging market state-owned enterprises

30. Sofie Blinkenberg Federspiel
IT, organisation og digitalisering:
Institutionelt arbejde i den kommunale digitaliseringsproces

31. Elvi Weinreich
Hvilke offentlige ledere er der brug for når velfærdstænkningen flytter sig – er Diplomuddannelsens lederprofil svaret?

32. Ellen Mølgaard Korsager
Self-conception and image of context in the growth of the firm – A Penrosian History of Fiberline Composites

33. Else Skjold
The Daily Selection

34. Marie Louise Conradsen
The Cancer Centre That Never Was
The Organisation of Danish Cancer Research 1949-1992

35. Virgilio Failla
Three Essays on the Dynamics of Entrepreneurs in the Labor Market

36. Nicky Nedergaard
Brand-Based Innovation
Relational Perspectives on Brand Logics and Design Innovation Strategies and Implementation

37. Mads Gjedsted Nielsen
Essays in Real Estate Finance

38. Kristin Martina Brandl
Process Perspectives on Service Offshoring

39. Mia Rosa Koss Hartmann
In the gray zone
With police in making space for creativity

40. Karen Ingerslev
Healthcare Innovation under The Microscope
Framing Boundaries of Wicked Problems

41. Tim Neerup Themsen
Risk Management in large Danish public capital investment programmes

2015

1. Jakob Ion Wille
Film som design
Design af levende billeder i film og tv-serier

2. Christiane Mossin
Interzones of Law and Metaphysics
Hierarchies, Logics and Foundations of Social Order seen through the Prism of EU Social Rights

3. Thomas Tøth
TRUSTWORTHINESS: ENABLING GLOBAL COLLABORATION
An Ethnographic Study of Trust, Distance, Control, Culture and Boundary Spanning within Offshore Outsourcing of IT Services

4. Steven Højlund
Evaluation Use in Evaluation Systems – The Case of the European Commission

5. Julia Kirch Kirkegaard
AMBIGUOUS WINDS OF CHANGE – OR FIGHTING AGAINST WINDMILLS IN CHINESE WIND POWER
A CONSTRUCTIVIST INQUIRY INTO CHINA'S PRAGMATICS OF GREEN MARKETISATION MAPPING CONTROVERSIES OVER A POTENTIAL TURN TO QUALITY IN CHINESE WIND POWER

6. Michelle Carol Antero
A Multi-case Analysis of the Development of Enterprise Resource Planning Systems (ERP) Business Practices

Morten Friis-Olivarius
The Associative Nature of Creativity

7. Mathew Abraham
New Cooperativism:
A study of emerging producer organisations in India

8. Stine Hedegaard
Sustainability-Focused Identity: Identity work performed to manage, negotiate and resolve barriers and tensions that arise in the process of constructing or ganizational identity in a sustainability context

9. Cecilie Glerup
Organizing Science in Society – the conduct and justification of resposible research

10. Allan Salling Pedersen
Implementering af ITIL® IT-governance - når best practice konflikter med kulturen Løsning af implementeringsproblemer gennem anvendelse af kendte CSF i et aktionsforskningsforløb.

11. Nihat Misir
A Real Options Approach to Determining Power Prices

12. Mamdouh Medhat
MEASURING AND PRICING THE RISK OF CORPORATE FAILURES

13. Rina Hansen
Toward a Digital Strategy for Omnichannel Retailing

14. Eva Pallesen
In the rhythm of welfare creation
A relational processual investigation moving beyond the conceptual horizon of welfare management

15. Gouya Harirchi
In Search of Opportunities: Three Essays on Global Linkages for Innovation

16. Lotte Holck
Embedded Diversity: A critical ethnographic study of the structural tensions of organizing diversity

17. Jose Daniel Balarezo
Learning through Scenario Planning

18. Louise Pram Nielsen
Knowledge dissemination based on terminological ontologies. Using eye tracking to further user interface design.

19. Sofie Dam
PUBLIC-PRIVATE PARTNERSHIPS FOR INNOVATION AND SUSTAINABILITY TRANSFORMATION
An embedded, comparative case study of municipal waste management in England and Denmark

20. Ulrik Hartmyer Christiansen
Follwoing the Content of Reported Risk Across the Organization

21. Guro Refsum Sanden
Language strategies in multinational corporations. A cross-sector study of financial service companies and manufacturing companies.

22. Linn Gevoll
Designing performance management for operational level
- A closer look on the role of design choices in framing coordination and motivation

23. Frederik Larsen
Objects and Social Actions
– on Second-hand Valuation Practices

24. Thorhildur Hansdottir Jetzek
The Sustainable Value of Open Government Data
Uncovering the Generative Mechanisms of Open Data through a Mixed Methods Approach

25. Gustav Toppenberg
Innovation-based M&A
– Technological-Integration Challenges – The Case of Digital-Technology Companies

26. Mie Plotnikof
Challenges of Collaborative Governance
An Organizational Discourse Study of Public Managers' Struggles with Collaboration across the Daycare Area

27. Christian Garmann Johnsen
Who Are the Post-Bureaucrats?
A Philosophical Examination of the Creative Manager, the Authentic Leader and the Entrepreneur

28. Jacob Brogaard-Kay
Constituting Performance Management
A field study of a pharmaceutical company

29. Rasmus Ploug Jenle
Engineering Markets for Control: Integrating Wind Power into the Danish Electricity System

30. Morten Lindholst
Complex Business Negotiation: Understanding Preparation and Planning

31. Morten Grynings
TRUST AND TRANSPARENCY FROM AN ALIGNMENT PERSPECTIVE

32. Peter Andreas Norn
Byregimer og styringsevne: Politisk lederskab af store byudviklingsprojekter

33. Milan Miric
Essays on Competition, Innovation and Firm Strategy in Digital Markets

34. Sanne K. Hjordrup
The Value of Talent Management
Rethinking practice, problems and possibilities

35. Johanna Sax
Strategic Risk Management
– Analyzing Antecedents and Contingencies for Value Creation

36. Pernille Rydén
Strategic Cognition of Social Media

37. Mimmi Sjöklint
The Measurable Me
- The Influence of Self-tracking on the User Experience

38. Juan Ignacio Staricco
Towards a Fair Global Economic Regime? A critical assessment of Fair Trade through the examination of the Argentinean wine industry

39. Marie Henriette Madsen
Emerging and temporary connections in Quality work

40. Yangfeng CAO
Toward a Process Framework of Business Model Innovation in the Global Context
Entrepreneurship-Enabled Dynamic Capability of Medium-Sized Multinational Enterprises

41. Carsten Scheibye
Enactment of the Organizational Cost Structure in Value Chain Configuration
A Contribution to Strategic Cost Management

2016

1. Signe Sofie Dyrby
Enterprise Social Media at Work

2. Dorte Boesby Dahl
The making of the public parking attendant
Dirt, aesthetics and inclusion in public service work

3. Verena Girschik
Realizing Corporate Responsibility Positioning and Framing in Nascent Institutional Change

4. Anders Ørding Olsen
IN SEARCH OF SOLUTIONS
Inertia, Knowledge Sources and Diversity in Collaborative Problem-solving

5. Pernille Steen Pedersen
Udkast til et nyt copingbegreb
En kvalifikation af ledelsesmuligheder for at forebygge sygefravær ved psykiske problemer.

6. Kerli Kant Hvass
Weaving a Path from Waste to Value: Exploring fashion industry business models and the circular economy

7. Kasper Lindskow
Exploring Digital News Publishing Business Models – a production network approach

8. Mikkel Mouritz Marfelt
The chameleon workforce:
Assembling and negotiating the content of a workforce

9. Marianne Bertelsen
Aesthetic encounters
Rethinking autonomy, space & time in today's world of art

10. Louise Hauberg Wilhelmsen
EU PERSPECTIVES ON INTERNATIONAL COMMERCIAL ARBITRATION

11. Abid Hussain
On the Design, Development and Use of the Social Data Analytics Tool (SODATO): Design Propositions, Patterns, and Principles for Big Social Data Analytics

12. Mark Bruun
Essays on Earnings Predictability

13. Tor Bøe-Lillegraven
BUSINESS PARADOXES, BLACK BOXES, AND BIG DATA: BEYOND ORGANIZATIONAL AMBIDEXTERITY

14. Hadis Khonsary-Atighi
ECONOMIC DETERMINANTS OF DOMESTIC INVESTMENT IN AN OIL-BASED ECONOMY: THE CASE OF IRAN (1965-2010)

15. Maj Lervad Grasten
Rule of Law or Rule by Lawyers?
On the Politics of Translation in Global Governance

16. Lene Granzau Juel-Jacobsen
SUPERMARKEDETS MODUS OPERANDI – en hverdagssociologisk undersøgelse af forholdet mellem rum og handlen og understøtte relationsopbygning?

17. Christine Thalsgård Henriques
In search of entrepreneurial learning – Towards a relational perspective on incubating practices?

18. Patrick Bennett
Essays in Education, Crime, and Job Displacement

19. Søren Korsgaard
Payments and Central Bank Policy

20. Marie Kruse Skibsted
Empirical Essays in Economics of Education and Labor

21. Elizabeth Benedict Christensen
The Constantly Contingent Sense of Belonging of the 1.5 Generation Undocumented Youth
An Everyday Perspective

22. Lasse J. Jessen
Essays on Discounting Behavior and Gambling Behavior

23. Kalle Johannes Rose
Når stifterviljen dør…
Et retsøkonomisk bidrag til 200 års juridisk konflikt om ejendomsretten

24. Andreas Søeborg Kirkedal
Danish Stød and Automatic Speech Recognition

25. Ida Lunde Jørgensen
Institutions and Legitimations in Finance for the Arts

26. Olga Rykov Ibsen
An empirical cross-linguistic study of directives: A semiotic approach to the sentence forms chosen by British, Danish and Russian speakers in native and ELF contexts

27. Desi Volker
Understanding Interest Rate Volatility

28. Angeli Elizabeth Weller
Practice at the Boundaries of Business Ethics & Corporate Social Responsibility

29. Ida Danneskiold-Samsøe
Levende læring i kunstneriske organisationer
En undersøgelse af læringsprocesser mellem projekt og organisation på Aarhus Teater

30. Leif Christensen
Quality of information – The role of internal controls and materiality

31. Olga Zarzecka
Tie Content in Professional Networks

32. Henrik Mahncke
De store gaver
- Filantropiens gensidighedsrelationer i teori og praksis

33. Carsten Lund Pedersen
Using the Collective Wisdom of Frontline Employees in Strategic Issue Management

34. Yun Liu
Essays on Market Design

35. Denitsa Hazarbassanova Blagoeva
The Internationalisation of Service Firms

36. Manya Jaura Lind
Capability development in an off-shoring context: How, why and by whom

37. Luis R. Boscán F.
Essays on the Design of Contracts and Markets for Power System Flexibility

38. Andreas Philipp Distel
Capabilities for Strategic Adaptation: Micro-Foundations, Organizational Conditions, and Performance Implications

39. Lavinia Bleoca
The Usefulness of Innovation and Intellectual Capital in Business Performance: The Financial Effects of Knowledge Management vs. Disclosure

40. Henrik Jensen
Economic Organization and Imperfect Managerial Knowledge: A Study of the Role of Managerial Meta-Knowledge in the Management of Distributed Knowledge

41. Stine Mosekjær
The Understanding of English Emotion Words by Chinese and Japanese Speakers of English as a Lingua Franca An Empirical Study

42. Hallur Tor Sigurdarson
The Ministry of Desire - Anxiety and entrepreneurship in a bureaucracy

43. Kätlin Pulk
Making Time While Being in Time
A study of the temporality of organizational processes

44. Valeria Giacomin
Contextualizing the cluster Palm oil in Southeast Asia in global perspective (1880s–1970s)

45. Jeanette Willert
Managers' use of multiple Management Control Systems: The role and interplay of management control systems and company performance

46. Mads Vestergaard Jensen
Financial Frictions: Implications for Early Option Exercise and Realized Volatility

47. Mikael Reimer Jensen
Interbank Markets and Frictions

48. Benjamin Faigen
Essays on Employee Ownership

49. Adela Michea
Enacting Business Models
An Ethnographic Study of an Emerging Business Model Innovation within the Frame of a Manufacturing Company.

50. Iben Sandal Stjerne
Transcending organization in temporary systems
Aesthetics' organizing work and employment in Creative Industries

51. Simon Krogh
Anticipating Organizational Change

52. Sarah Netter
Exploring the Sharing Economy

53. Lene Tolstrup Christensen
State-owned enterprises as institutional market actors in the marketization of public service provision:
A comparative case study of Danish and Swedish passenger rail 1990–2015

54. Kyoung(Kay) Sun Park
Three Essays on Financial Economics

2017

1. Mari Bjerck
Apparel at work. Work uniforms and women in male-dominated manual occupations.

2. Christoph H. Flöthmann
Who Manages Our Supply Chains? Backgrounds, Competencies and Contributions of Human Resources in Supply Chain Management

3. Aleksandra Anna Rzeźnik
Essays in Empirical Asset Pricing

4. Claes Bäckman
Essays on Housing Markets

5. Kirsti Reitan Andersen
Stabilizing Sustainability
in the Textile and Fashion Industry

6. Kira Hoffmann
Cost Behavior: An Empirical Analysis of Determinants and Consequences of Asymmetries

7. Tobin Hanspal
Essays in Household Finance

8. Nina Lange
Correlation in Energy Markets

9. Anjum Fayyaz
Donor Interventions and SME Networking in Industrial Clusters in Punjab Province, Pakistan

10. Magnus Paulsen Hansen
Trying the unemployed. Justification and critique, emancipation and coercion towards the 'active society'. A study of contemporary reforms in France and Denmark

11. Sameer Azizi
Corporate Social Responsibility in Afghanistan
– a critical case study of the mobile telecommunications industry

12. Malene Myhre
The internationalization of small and medium-sized enterprises:
A qualitative study

13. Thomas Presskorn-Thygesen
The Significance of Normativity – Studies in Post-Kantian Philosophy and Social Theory

14. Federico Clementi
Essays on multinational production and international trade

15. Lara Anne Hale
Experimental Standards in Sustainability Transitions: Insights from the Building Sector

16. Richard Pucci
Accounting for Financial Instruments in an Uncertain World
Controversies in IFRS in the Aftermath of the 2008 Financial Crisis

17. Sarah Maria Denta
Kommunale offentlige private partnerskaber
Regulering I skyggen af Farumsagen

18. Christian Östlund
Design for e-training

19. Amalie Martinus Hauge
Organizing Valuations – a pragmatic inquiry

20. Tim Holst Celik
Tension-filled Governance? Exploring the Emergence, Consolidation and Reconfiguration of Legitimatory and Fiscal State-crafting

21. Christian Bason
Leading Public Design: How managers engage with design to transform public governance

22. Davide Tomio
Essays on Arbitrage and Market Liquidity

23. Simone Stæhr
Financial Analysts' Forecasts
Behavioral Aspects and the Impact of Personal Characteristics

24. Mikkel Godt Gregersen
Management Control, Intrinsic Motivation and Creativity
– How Can They Coexist

25. Kristjan Johannes Suse Jespersen
Advancing the Payments for Ecosystem Service Discourse Through Institutional Theory

26. Kristian Bondo Hansen
Crowds and Speculation: A study of crowd phenomena in the U.S. financial markets 1890 to 1940

27. Lars Balslev
Actors and practices – An institutional study on management accounting change in Air Greenland

28. Sven Klingler
Essays on Asset Pricing with Financial Frictions

29. Klement Ahrensbach Rasmussen
Business Model Innovation
The Role of Organizational Design

30. Giulio Zichella
Entrepreneurial Cognition.
Three essays on entrepreneurial behavior and cognition under risk and uncertainty

31. Richard Ledborg Hansen
En forkærlighed til det eksisterende – mellemlederens oplevelse af forandringsmodstand i organisatoriske forandringer

32. Vilhelm Stefan Holsting
Militært chefvirke: Kritik og retfærdiggørelse mellem politik og profession

33. Thomas Jensen
Shipping Information Pipeline: An information infrastructure to improve international containerized shipping

34. Dzmitry Bartalevich
Do economic theories inform policy? Analysis of the influence of the Chicago School on European Union competition policy

35. Kristian Roed Nielsen
Crowdfunding for Sustainability: A study on the potential of reward-based crowdfunding in supporting sustainable entrepreneurship

36. Emil Husted
There is always an alternative: A study of control and commitment in political organization

37. Anders Ludvig Sevelsted
Interpreting Bonds and Boundaries of Obligation. A genealogy of the emergence and development of Protestant voluntary social work in Denmark as shown through the cases of the Copenhagen Home Mission and the Blue Cross (1850 – 1950)

38. Niklas Kohl
Essays on Stock Issuance

39. Maya Christiane Flensborg Jensen
BOUNDARIES OF PROFESSIONALIZATION AT WORK
An ethnography-inspired study of care workers' dilemmas at the margin

40. Andreas Kamstrup
Crowdsourcing and the Architectural Competition as Organisational Technologies

41. Louise Lyngfeldt Gorm Hansen
Triggering Earthquakes in Science, Politics and Chinese Hydropower - A Controversy Study

2018

1. Vishv Priya Kohli
Combatting Falsification and Counterfeiting of Medicinal Products in the European Union – A Legal Analysis

2. Helle Haurum
Customer Engagement Behavior in the context of Continuous Service Relationships

3. Nis Grünberg
The Party-state order: Essays on China's political organization and political economic institutions

4. Jesper Christensen
A Behavioral Theory of Human Capital Integration

5. Poula Marie Helth
Learning in practice

6. Rasmus Vendler Toft-Kehler
Entrepreneurship as a career? An investigation of the relationship between entrepreneurial experience and entrepreneurial outcome

7. Szymon Furtak
Sensing the Future: Designing sensor-based predictive information systems for forecasting spare part demand for diesel engines

8. Mette Brehm Johansen
Organizing patient involvement. An ethnographic study

9. Iwona Sulinska
Complexities of Social Capital in Boards of Directors

10. Cecilie Fanøe Petersen
Award of public contracts as a means to conferring State aid: A legal analysis of the interface between public procurement law and State aid law

11. Ahmad Ahmad Barirani
Three Experimental Studies on Entrepreneurship

12. Carsten Allerslev Olsen
Financial Reporting Enforcement: Impact and Consequences

13. Irene Christensen
New product fumbles – Organizing for the Ramp-up process

14. Jacob Taarup-Esbensen
Managing communities – Mining MNEs' community risk management practices

15. Lester Allan Lasrado
Set-Theoretic approach to maturity models

16. Mia B. Münster
Intention vs. Perception of Designed Atmospheres in Fashion Stores

17. Anne Sluhan
Non-Financial Dimensions of Family Firm Ownership: How Socioemotional Wealth and Familiness Influence Internationalization

18. Henrik Yde Andersen
Essays on Debt and Pensions

19. Fabian Heinrich Müller
Valuation Reversed – When Valuators are Valuated. An Analysis of the Perception of and Reaction to Reviewers in Fine-Dining

20. Martin Jarmatz
Organizing for Pricing

21. Niels Joachim Christfort Gormsen
Essays on Empirical Asset Pricing

22. Diego Zunino
Socio-Cognitive Perspectives in Business Venturing

23. Benjamin Asmussen
Networks and Faces between Copenhagen and Canton, 1730-1840

24. Dalia Bagdziunaite
Brains at Brand Touchpoints
A Consumer Neuroscience Study of Information Processing of Brand Advertisements and the Store Environment in Compulsive Buying

25. Erol Kazan
Towards a Disruptive Digital Platform Model

26. Andreas Bang Nielsen
Essays on Foreign Exchange and Credit Risk

27. Anne Krebs
Accountable, Operable Knowledge Toward Value Representations of Individual Knowledge in Accounting

28. Matilde Fogh Kirkegaard
A firm- and demand-side perspective on behavioral strategy for value creation: Insights from the hearing aid industry

29. Agnieszka Nowinska
SHIPS AND RELATION-SHIPS
Tie formation in the sector of shipping intermediaries in shipping

30. Stine Evald Bentsen
The Comprehension of English Texts by Native Speakers of English and Japanese, Chinese and Russian Speakers of English as a Lingua Franca. An Empirical Study.

31. Stine Louise Daetz
Essays on Financial Frictions in Lending Markets

32. *Christian Skov Jensen*
Essays on Asset Pricing

33. Anders Kryger
Aligning future employee action and corporate strategy in a resource-scarce environment

34. Maitane Elorriaga-Rubio
The behavioral foundations of strategic decision-making: A contextual perspective

35. Roddy Walker
Leadership Development as Organisational Rehabilitation: Shaping Middle-Managers as Double Agents

36. Jinsun Bae
Producing Garments for Global Markets Corporate social responsibility (CSR) in Myanmar's export garment industry 2011–2015

37. Queralt Prat-i-Pubill
Axiological knowledge in a knowledge driven world. Considerations for organizations.

38. Pia Mølgaard
Essays on Corporate Loans and Credit Risk

39. Marzia Aricò
Service Design as a Transformative Force: Introduction and Adoption in an Organizational Context

40. Christian Dyrlund Wåhlin-Jacobsen
Constructing change initiatives in workplace voice activities Studies from a social interaction perspective

41. Peter Kalum Schou
Institutional Logics in Entrepreneurial Ventures: How Competing Logics arise and shape organizational processes and outcomes during scale-up

42. Per Henriksen
Enterprise Risk Management Rationaler og paradokser i en moderne ledelsesteknologi

43. Maximilian Schellmann
The Politics of Organizing Refugee Camps

44. Jacob Halvas Bjerre
Excluding the Jews: The Aryanization of Danish-German Trade and German Anti-Jewish Policy in Denmark 1937-1943

45. Ida Schrøder
Hybridising accounting and caring: A symmetrical study of how costs and needs are connected in Danish child protection work

46. Katrine Kunst
Electronic Word of Behavior: Transforming digital traces of consumer behaviors into communicative content in product design

47. Viktor Avlonitis
Essays on the role of modularity in management: Towards a unified perspective of modular and integral design

48. Anne Sofie Fischer
Negotiating Spaces of Everyday Politics:
-An ethnographic study of organizing for social transformation for women in urban poverty, Delhi, India

2019

1. Shihan Du
   ESSAYS IN EMPIRICAL STUDIES BASED ON ADMINISTRATIVE LABOUR MARKET DATA

2. Mart Laatsit
   Policy learning in innovation policy: A comparative analysis of European Union member states

3. Peter J. Wynne
   Proactively Building Capabilities for the Post-Acquisition Integration of Information Systems

4. Kalina S. Staykova
   Generative Mechanisms for Digital Platform Ecosystem Evolution

5. Ieva Linkeviciute
   Essays on the Demand-Side Management in Electricity Markets

6. Jonatan Echebarria Fernández
   Jurisdiction and Arbitration Agreements in Contracts for the Carriage of Goods by Sea – Limitations on Party Autonomy

7. Louise Thorn Bøttkjær
   Votes for sale. Essays on clientelism in new democracies.

8. Ditte Vilstrup Holm
   The Poetics of Participation: the organizing of participation in contemporary art

9. Philip Rosenbaum
   Essays in Labor Markets – Gender, Fertility and Education

10. Mia Olsen
    Mobile Betalinger - Succesfaktorer og Adfærdsmæssige Konsekvenser

11. Adrián Luis Mérida Gutiérrez
    Entrepreneurial Careers: Determinants, Trajectories, and Outcomes

12. Frederik Regli
    Essays on Crude Oil Tanker Markets

13. Cancan Wang
    Becoming Adaptive through Social Media: Transforming Governance and Organizational Form in Collaborative E-government

14. Lena Lindbjerg Sperling
    Economic and Cultural Development: Empirical Studies of Micro-level Data

15. Xia Zhang
    Obligation, face and facework: An empirical study of the communicative act of cancellation of an obligation by Chinese, Danish and British business professionals in both L1 and ELF contexts

16. Stefan Kirkegaard Sløk-Madsen
    Entrepreneurial Judgment and Commercialization

17. Erin Leitheiser
    The Comparative Dynamics of Private Governance
    The case of the Bangladesh Ready-Made Garment Industry

18. Lone Christensen
    STRATEGIIMPLEMENTERING: STYRINGSBESTRÆBELSER, IDENTITET OG AFFEKT

19. Thomas Kjær Poulsen
    Essays on Asset Pricing with Financial Frictions

20. Maria Lundberg
    Trust and self-trust in leadership identity constructions: A qualitative exploration of narrative ecology in the discursive aftermath of heroic discourse

21. Tina Joanes
*Sufficiency for sustainability
Determinants and strategies for reducing clothing consumption*

22. Benjamin Johannes Flesch
Social Set Visualizer (SoSeVi): Design, Development and Evaluation of a Visual Analytics Tool for Computational Set Analysis of Big Social Data

23. Henriette Sophia Groskopff Tvede Schleimann
Creating innovation through collaboration – Partnering in the maritime sector

24. Kristian Steensen Nielsen
The Role of Self-Regulation in Environmental Behavior Change

25. Lydia L. Jørgensen
Moving Organizational Atmospheres

26. Theodor Lucian Vladasel
Embracing Heterogeneity: Essays in Entrepreneurship and Human Capital

27. Seidi Suurmets
Contextual Effects in Consumer Research: An Investigation of Consumer Information Processing and Behavior via the Applicati on of Eye-tracking Methodology

28. Marie Sundby Palle Nickelsen
Reformer mellem integritet og innovation: Reform af reformens form i den danske centraladministration fra 1920 til 2019

29. Vibeke Kristine Scheller
The temporal organizing of same-day discharge: A tempography of a Cardiac Day Unit

30. Qian Sun
Adopting Artificial Intelligence in Healthcare in the Digital Age: Perceived Challenges, Frame Incongruence, and Social Power

31. Dorthe Thorning Mejlhede
Artful change agency and organizing for innovation – the case of a Nordic fintech cooperative

32. Benjamin Christoffersen
*Corporate Default Models:
Empirical Evidence and Methodological Contributions*

TITLER I ATV PH.D.-SERIEN

1992
1. Niels Kornum
Servicesamkørsel – organisation, økonomi og planlægningsmetode

1995
2. Verner Worm
Nordiske virksomheder i Kina
Kulturspecifikke interaktionsrelationer ved nordiske virksomhedsetableringer i Kina

1999
3. Mogens Bjerre
Key Account Management of Complex Strategic Relationships
An Empirical Study of the Fast Moving Consumer Goods Industry

2000
4. Lotte Darsø
Innovation in the Making
Interaction Research with heterogeneous Groups of Knowledge Workers creating new Knowledge and new Leads

2001
5. Peter Hobolt Jensen
Managing Strategic Design Identities
The case of the Lego Developer Network

2002
6. Peter Lohmann
The Deleuzian Other of Organizational Change – Moving Perspectives of the Human

7. Anne Marie Jess Hansen
To lead from a distance: The dynamic interplay between strategy and strategizing – A case study of the strategic management process

2003
8. Lotte Henriksen
Videndeling
– om organisatoriske og ledelsesmæssige udfordringer ved videndeling i praksis

9. Niels Christian Nickelsen
Arrangements of Knowing: Coordinating Procedures Tools and Bodies in Industrial Production – a case study of the collective making of new products

2005
10. Carsten Ørts Hansen
Konstruktion af ledelsesteknologier og effektivitet

TITLER I DBA PH.D.-SERIEN

2007
1. Peter Kastrup-Misir
Endeavoring to Understand Market Orientation – and the concomitant co-mutation of the researched, the re searcher, the research itself and the truth

2009
1. Torkild Leo Thellefsen
Fundamental Signs and Significance effects
A Semeiotic outline of Fundamental Signs, Significance-effects, Knowledge Profiling and their use in Knowledge Organization and Branding

2. Daniel Ronzani
When Bits Learn to Walk Don’t Make Them Trip. Technological Innovation and the Role of Regulation by Law in Information Systems Research: the Case of Radio Frequency Identification (RFID)

2010
1. Alexander Carnera
Magten over livet og livet som magt
Studier i den biopolitiske ambivalens