CARNEGIE MELLON UNIVERSITY

LONGITUDINAL MIXED MEMBERSHIP MODELS WITH APPLICATIONS TO DISABILITY SURVEY DATA

A DISSERTATION SUBMITTED TO THE GRADUATE SCHOOL IN PARTIAL
FULFILLMENT OF THE REQUIREMENTS

for the degree

DOCTOR OF PHILOSOPHY
In
STATISTICS

by

DANIEL MANRIQUE-VALLIER

Department of Statistics
Carnegie Mellon University
Pittsburgh, Pennsylvania 15213

August, 2010

Abstract

When analyzing longitudinal data we need to balance our understanding of individual variability with the production of meaningful and interpretable summaries of overall population tendencies. This is especially true when those in the target population are known to be heterogeneous in their progression over time due to unobserved individual traits. Additional complications arise when the data are discrete and multivariate. I propose a new family of models to analyze such data by combining features from a version of the cross-sectional Grade of Membership Model (Woodbury et al., 1978; Erosheva et al., 2007) and from the longitudinal Multivariate Latent Trajectory Model (Connor, 2006). This new family of models works by considering individuals to be combinations of a small number of "ideal" or "extreme" classes. By describing the ways each of these extreme classes evolves over time we are able to describe distinct general tendencies. At the same time, by considering individuals to be individual-level mixtures of these profiles, we are able to handle complex forms of heterogeneity.

I apply my method to data from the National Long Term Care Survey (NLTCS), a longitudinal survey with six completed waves aimed to assess the state and characteristics of disability among U.S. citizens age 65 and above. I develop a full Bayesian specification and estimation methods based on Markov Chain Monte Carlo sampling, as well as several extensions to the basic methods to handle specific aspects of the NLTCS and to be able to answer relevant questions about the differences between disability patterns across different generations and the interplay between disability dynamics and survival times.

Acknowledgements

I want to express my gratitude to my advisor, Stephen Fienberg. More than a thesis advisor, he has been a true mentor. Through his encouragement, enthusiastic support, sound advice, gentle pressure and infinite patience, he has not only made this thesis possible, but also helped me grow in many ways. I realize how fortunate I am to have had him as my advisor and look forward to future collaborations.

Many thanks to the rest of my thesis committee. Brian Junker, Howard Seltman and Elena Erosheva have all provided support, valuable suggestions and well targeted criticism that have greatly improved the quality of this document. Likewise, I want to thank Professors Bill Eddy and Jay Kadane for teaching me, from very different perspectives, how to face the challenges of statistical research.

Many friends and classmates have made my time at CMU an enriching and pleasurable experience. Justin Gross, Zhanwu Liu, Christian Niyonkuru, Cindy Yang, Eunice Kim, Judy Xi and Linqiao Zhao have always been there for me. Susan Buchman and Han Liu made that year in PH117 one of the best educational experiences I have ever had. Avranil Sarkar, Gaia Bellone, Estefa Etchegaray and April Galyardt made my writing years smooth and pleasurable.

Thanks to my parents, Natty and Nelson, who have supported—and tolerated—me throughout all these years of trying to find my path in life.

Finally, my deepest thanks go my wife, Lucia, for her loving and unconditional support. From graciously tolerating random interruptions ("look at my pretty graph!") to painstakingly proofreading several manuscripts, she has also had a direct involvement in this thesis.

I realize I could not have done it without her.

I dedicate this thesis to Lucia and to our sons, Adrián and Mateo. They have shown me that there is more in life than I had ever imagined.

Contents

Abstract **iv**

Acknowledgements **v**

1 Introduction **1**

1.1 Data: The National Long Term Care Survey 2

1.2 Notation and structure 7

1.3 Modeling Functional Disability in the NLTCS: An Overview 9

1.3.1 Mixed membership models. The Grade of Membership Model 12

1.3.2 Modeling Longitudinal Patterns - Latent Trajectory Models 16

1.4 Overview of the thesis 18

2 Basic Grade of Membership Trajectory Model **20**

2.1 General Formulation 22

2.1.1 What To Do With The Missing Values? 25

2.2 Continuous Trajectories GoM Model 27

2.2.1 Estimation via Markov Chain Monte Carlo 29

2.3 A Variant: Discrete Piecewise Constant Trajectories 37

2.3.1 Estimation 39

2.4 Comments 40

2.A Appendices 43

2.A.1 Computational Issues 43

3 Applying the Basic Model to the NLTCS data 48
3.1 Preparation and Preliminary Issues 48
  3.1.1 Data 48
  3.1.2 Prior distributions 49
  3.1.3 Execution of the Algorithm 51
3.2 Results 52
3.3 Using the Discrete Trajectories Variant 57
3.4 Discussion 73
3.A Appendix: Non-negative Priors for Trajectory Slope Parameters 75

4 Modeling Generational Differences: Grouped Data Model 78
4.1 Introduction 78
4.2 Cohorts as covariates 80
4.3 Extending the Basic Model to Handle Generational Groups 81
4.4 Estimation 84
4.5 Further Remarks 86

5 Applying the Grouped Model to the NLTCS 88
5.1 Preliminaries 89
5.2 Results 91
5.3 Discussion 94
5.A Appendix: Posterior Estimates of Parameters 97

6 Modeling Mortality 103
6.1 Modeling Mortality on the NLTCS 105
6.2 Practical Aspects of Modeling Mortality for NLTCS subjects 107
  6.2.1 Initial Considerations 107
  6.2.2 Dealing with Right Censoring 108
  6.2.3 Dealing with left truncation 109
6.3 A Weibull model With Truncation and Delayed Entry 110
  6.3.1 Estimation Using MCMC 113

6.3.2 Testing the Weibull survival model with NLTCS mortality data . . . 116

6.4 Constructing a Joint Model for Mortality and Disability 118

6.4.1 A GoM version of the simple Weibull survival model 118

6.4.2 Combining the Basic GoM Trajectory Model and Weibull Mortality Model 121

6.4.3 Detailed Implementation 124

6.4.4 Missing Values 125

6.4.5 Posterior Estimation Using MCMC 129

6.5 Summary and Discussion 132

7 Applying the Joint Disability-Mortality model to the NLTCS Data 137

7.1 Preliminaries 138

7.2 Results 139

7.3 Evaluating the Survival Truncation Approximation 148

7.4 Understanding the Trajectory Profiles: Using Step Trajectories with the Joint Model 150

7.5 Discussion 154

7.A Appendix - Posterior Summaries 155

8 Summary and Future Work 159

8.1 Contributions 160

8.2 Directions for Future Work 163

List of Figures

1.1 Simplified diagram of NLTCS interview flow. 5

2.1 "Realignment effect" when translating the discrete measurements taken from different individuals to a common age scale. 43

3.1 Prior and posterior distributions for parameters α_0 and β_{021} for basic model with $K = 3$. The prior distribution for α_0 is Gamma$(1, 5)$ and for β_{0jk} is $N(0, 100)$. 54

3.2 Extreme trajectories of disability over time for each ADL and extreme profile ($K = 2$) Dashed lines indicate the point where the probability of acquiring a disability in ADL j reaches 50%. 59

3.3 Extreme trajectories of disability over time for each ADL and extreme profile ($K = 3$) Dashed lines indicate the point where the probability of acquiring a disability in ADL j reaches 50%. 61

3.4 Extreme trajectories of disability over time for each ADL and extreme profile ($K = 4$) Dashed lines indicate the point where the probability of acquiring a disability in ADL j reaches 50%. 63

3.5 Extreme trajectories of disability over time for each ADL and extreme profile ($K = 5$) Dashed lines indicate the point where the probability of acquiring a disability in ADL j reaches 50%. 65

3.6 Individual-level mixture of trajectories for model with $K = 2$ extreme profiles for each ADL. Extreme trajectories are represented with thick lines and and a random sample of 100 individual posterior trajectory curves are plotted using thin lines 67

3.7 Individual-level mixture of trajectories for model with $K = 3$ extreme profiles for each ADL. Extreme trajectories are represented with thick lines and and a random sample of 100 individual posterior trajectory curves are plotted using thin lines 68

3.8 Individual-level mixture of trajectories for model with $K = 4$ extreme profiles for each ADL. Extreme trajectories are represented with thick lines and and a random sample of 100 individual posterior trajectory curves are plotted using thin lines 69

3.9 Extreme trajectories of disability over time for each ADL and extreme profile ($K = 3$) for models with piecewise constant and with continuous trajectories. 70

3.10 Extreme trajectories of disability over time for each ADL and extreme profile ($K = 4$) for models with piecewise constant and continuous trajectories. 71

3.11 Extreme trajectories of disability over time for each ADL and extreme profile ($K = 5$) for models with piecewise constant and continuous trajectories. 72

5.1 Distribution of age coverage for all defined generational groups. 90

5.2 Evolution of the parameter vector ξ across different generations for model with $K = 2$ extreme profiles. The error bars show the 95% equal tail posterior credible intervals associated with the k-th component of the vector ξ. 93

5.3 Evolution of the parameter vector ξ across different generations for model with $K = 3$ extreme profiles. The error bars show the 95% equal tail posterior credible intervals associated with the k-th component of the vector ξ. 93

5.4 Evolution of the parameter vector ξ across different generations for model with $K = 4$ extreme profiles. The error bars show the 95% equal tail posterior credible intervals associated with the k-th component of the vector ξ. 94

5.5 Comparison between the extreme trajectories for ADL-5 for a model with $K = 3$ extreme profiles and the distribution of age coverage for generations 1 (born before 1906) and 5 (born after 1926). Although by design none of these cohorts spans the whole interval of $65 - 90$ years, both include regions where the profiles are well separated. 96

5.6 Comparison between the extreme trajectories for ADL-5 for a model with $K = 4$ extreme profiles and the distribution of age coverage for generations 1 (born before 1906) and 5 (born after 1926). Different from the case of $K = 3$, Generation 1 does not span a region where profiles $k = 1$ and $k = 2$ are well separated. 97

6.1 Density of the Weibull distribution for different values of the shape parameter α ($\beta = 1$). 112

6.2 Estimated Weibull survival curve for NLTCS subjects compared with no deleyed-entry corrected Kaplan-Meier estimate (left panel) and with hazards based estimate with delayed entry correction (right panel). 117

7.1 Extreme profiles of survival probability and disability trajectories ($K = 2$). 142

7.2 Extreme profiles of survival probability and disability trajectories ($K = 3$). 143

7.3 Extreme profiles of survival probability and disability trajectories ($K = 4$). 144

7.4 Comparison between disability extreme trajectories from Joint (solid line) and Basic (dashed line) models with $K = 3$ extreme profiles. 145

7.5 Estimated population-level survival curve for the joint model with $K = 3$ extreme profiles, compared with a non-parametric hazards based estimate as reference. 147

7.6 True value of the truncated mixed membership survival density (Equation 6.9) evaluated at the posterior expected time of death Vs. surrogate approximation (Equation 6.10) for 1000 random individuals from the sample. Model fitted with $K = 3$ extreme profiles. 150

7.7 Extreme trajectories of the joint model with discrete trajectories specification. 153

List of Tables

1.1 The six ADLs used in the analysis. 4

3.1 Values of proposal standard deviation σ^*_α for basic model with continuous logistic trajectories. 51

3.2 Posterior means for parameters of interest for basic model with $K = 2$ extreme profiles. Numbers between parenthesis are posterior standard deviations. 58

3.3 Posterior means for parameters of interest for basic model with $K = 3$ extreme profiles. Numbers between parenthesis are posterior standard deviations. 60

3.4 Posterior means for parameters of interest for basic model with $K = 4$ extreme profiles. Numbers between parenthesis are posterior standard deviations. 62

3.5 Posterior means for parameters of interest for basic model with $K = 5$ extreme profiles. Numbers between parenthesis are posterior standard deviations. 64

3.6 Observed and mean predictive posterior aggregated counts for the 10 most populated response patterns by age range, for basic model with $K = 2, 3, 4, 5$. Within each age group the response patterns are sorted decreasingly according to observed frequencies. For each model the X^2 statistic has been computed from the complete $2^6 \times 4$ contingency table. 66

3.7 Posterior means for parameters α_0 and ξ for basic model with discrete step trajectories for $K = 2, 3, 4, 5$ extreme profiles. Numbers between brackets are posterior standard deviations. 73

5.1 Definition of cohorts and number of measurements at each wave. Cohort 5 only have measurements in the last three waves of the NLTCS. 89

5.2 Posterior means for parameters of interest for basic model with $K = 2$ extreme profiles. Numbers between parenthesis are posterior standard deviations. 98

5.3 Posterior means for parameters of interest for basic model with $K = 3$ extreme profiles. Numbers between parenthesis are posterior standard deviations. 99

5.4 Posterior means for parameters of interest for basic model with $K = 4$ extreme profiles. Numbers between parenthesis are posterior standard deviations. 100

5.5 Posterior means for parameters of interest for basic model with $K = 5$ extreme profiles. Numbers between parenthesis are posterior standard deviations. 101

5.6 Observed and mean predictive posterior aggregated counts for the 10 most populated response patterns by age range, for the Grouped model with $K = 2, 3, 4, 5$. Within each age group the response patterns are sorted decreasingly according to observed frequencies. For each model the X^2 statistic has been computed from the complete $2^6 \times 4$ contingency table. . . 102

7.1 Values of proposal standard deviation σ^*_α for joint survival-disability model. 139

7.2 Observed and mean predictive posterior aggregated counts for the 10 most populated response patterns by age range, for basic model with $K = 2, 3, 4, 5$. Within each age group the response patterns are sorted decreasingly according to observed frequencies. 146

7.3 Comparison between fit summary (X^2 statistic computed from posterior predictive disability response counts stratified by age categories) of Basic and Joint model for different values of K. 147

7.4 Posterior means for parameters of interest for survival model with $K = 2$ extreme profiles. Numbers between parenthesis are posterior standard deviations. 156

7.5 Posterior means for parameters of interest for survival model with $K = 3$ extreme profiles. Numbers between parenthesis are posterior standard deviations. 157

7.6 Posterior means for parameters of interest for survival model with $K = 4$ extreme profiles. Numbers between parenthesis are posterior standard deviations. 158

Chapter 1

Introduction

In this thesis I propose models and estimation procedures to deal with discrete multivariate longitudinal data obtained from a heterogeneous population. Data with these characteristics is frequent in large scale longitudinal and panel surveys, designed to measure a phenomenon that evolves over time. Techniques that characterize such processes have to take into account this longitudinal nature, consisting of multiple sequential measures from the same individuals. At the same time, when the population is known or suspected to be heterogeneous, it is important to identify and model this individual variability. Moreover, the characterization of this heterogeneity can be revealing in and of itself.

The specific methods proposed and developed here are motivated by the analysis of data from the National Long Term Care Survey (NLTCS). The NLTCS is a longitudinal panel survey instrument aimed to assess chronic disability among the elderly population in the United States. Through its analysis, researchers seek to answer important questions related to the aging process and the prevalence of disability in the U.S.: How many elder

Americans will live with disabilities? What is the of duration of disability episodes? What is the age of onset of disability? Is it changing for younger generations? (see e.g. Connor et al. 2006). Answers to these questions are of great importance in public policy design due to, among other reasons, the increased public and private expenditure for disabled people in contrast with their able peers (Manton et al., 1997).

Many of the relevant public policy questions for which the NLTCS can potentially provide answers have to do with changes over time: changes during the life of an individual ("how is this individual likely to age?") or comparing people across different generations ("are people from later generations acquiring disabilities in a different way different than people born 20 years before?"). To answer these questions we need to look at the data longitudinally, because the same individuals are measured at multiple points in time and because the questions of interest involve individual disability profiles over time. In addition, several studies (e.g. Connor 2006; Stallard 2005) have shown that elderly American people are a heterogeneous population, meaning that not everyone could be expected to age the same way. Thus models for longitudinal disability data need to be capable of representing such heterogeneity.

1.1 Data: The National Long Term Care Survey

The National Long Term Care Survey (NLTCS) is a longitudinal panel survey designed specifically to assess the state and progression of chronic disability among elderly population in the United States (Corder and Manton, 1991). Its target population consists of people aged 65 years and older that present functional limitations lasting or expected to last 90 or more days (White, 2008). It was funded by the National Institute of Aging, designed and overseen by the Center for Demographic Studies at Duke University and implemented by the U.S. Census Bureau.

The NLTCS consists of six waves conducted in 1982, 1984, 1989, 1994, 1999 and 2004. Each wave consists of interviews to about 20,000 people, from which about 15,000 are previously interviewed individuals and around 5,000 are new sampled subjects to replace those who have died since the previous wave, except in the first wave (Clark, 1998). In aggregate, a total of around 49,000 people have been screened in the survey between 1982 and 2004.

The sampling frame of the NLTCS is the Medicare file. Since 97% of U.S. seniors 65+ years old are registered in it, this list is believed to be a good representation of the elderly population of the U.S. (Corder and Manton, 1991). Despite its census-like nature, however, potential differences between the intended target population and the sampling frame have led some researchers to propose to consider it only as representative of the Medicare records (Cox, 2009).

The NLTCS was designed to be used as several cross-sectional samples, considering each wave as a different sample from the target population at that time, and also as a longitudinal sample, following individuals across different measurement waves.

The NLTCS approaches disability through the measurement of individuals' capacity to perform a set of "Activities of Daily Living" (ADL) such as eating, bathing or dressing and "Instrumental Activities of Daily Living" (IADL) such as preparing meals or maintaining finances. Broadly stated, ADLs seek to measure people's ability to take care of themselves at a fundamental level, while IADLs measure the ability of living independently within a community (Connor, 2006). The survey instruments register these variables as a series of

ADL (j)	Abrv	Description
1	EAT	Eating
2	DRS	Dressing
3	TLT	Toileting
4	BED	Getting in and out of bed
5	MOB	Inside mobility
6	BTH	Bathing

Table 1.1: The six ADLs used in the analysis.

answers to “triggering questions” that are then summarized into a set of binary responses according to a set of decision rules. These binary responses indicate the presence or absence of impairments to perform such activities. Table 1.1 shows the list of ADLs that I have used in my own analyses.

Every individual sampled in the survey is first “screened” using a special questionnaire aimed at quickly detecting if he or she is “chronically disabled” (presenting a disability in some ADL or IADL lasting or expected to last 90 or more days). If screened-out, their status is registered and they are rescreened in subsequent waves to assess if the disability status has changed. If the individual is screened-in, he or she is then interviewed using a “detailed questionnaire”. There are different detailed questionnaires for institutionalized and community dweller individuals. After receiving a detailed questionnaire for the first time, the subject is then eligible to receive detailed questionnaires in all subsequent waves of the survey until death, to check if their functional status has changed (Clark, 1998). Figure 1.1 shows a simplified flow diagram of the interview and questionnaire eligibility process.

There are some exceptions to these rules. The 1994, 1999 and 2004 waves introduced supplementary samples of people aged 95 years or older. This component, called the

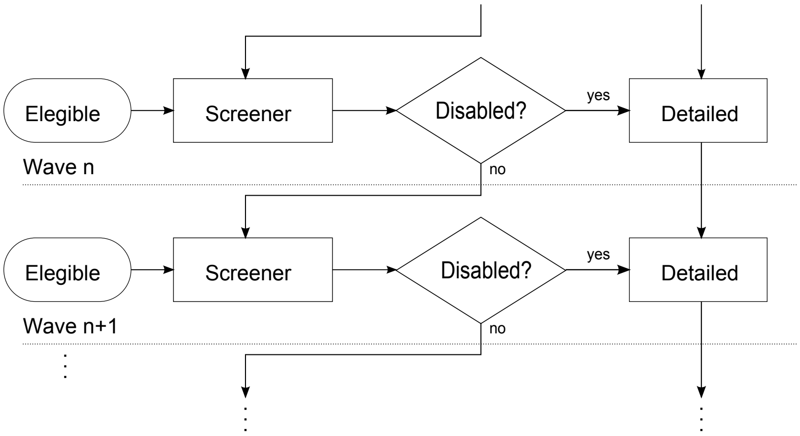


Figure 1.1: Simplified diagram of NLTCS interview flow.

"oldest old sample" was introduced to have more precision in the estimates of this group of people, since they tend to be scarce due to their high mortality rates (Stallard, 2005; White, 2008). The other main exception is a group of non-disabled individuals that, starting in 1994 were selected to receive the detailed interview, even though they would have been rejected according to the initial screening process (Stallard, 2005).

Another thing to bear in mind is that IADL information was only registered through the community detailed questionnaires. This means that IADL information is missing by design for institutionalized people, even if they were screened-in. Additionally, during the 1982 wave, where was no detailed questionnaire for institutionalized people, even if they screened-in.

There are some issues with the comparability between corresponding ADLs and IADLs between different instruments (screener and detailed interviews), as the triggering questions in each instrument are different and in some cases are oriented to measure different aspects of disability (Erosheva and White, 2009, 2010). One important aspect is that the screening instrument checks for at least one chronic limitation (lasting 90 or more days) while the detailed interviews do not. These issues could bring problems since people that were screened-in were automatically given the detailed forms in all subsequent waves and therefore assessed in potentially different variables, with the same label. Although this is an important issue that deserves further exploration, it falls beyond the scope of this thesis.

The response rate of the survey is extremely high, on the order of 95% for all waves, both because of the use of highly trained interviewers and the allowance of proxy respondents (Connor, 2006). Although the reliability of proxy response is a source of legitimate concern, I also consider this and related issues beyond the scope of the present work.

Although the NLTCS instruments try to track mortality, the reliability of these records is dubious, due to the difficulty in obtaining this information by just attempting followup contact. For this reason, researchers at Duke University undertook a special project to link the NLTCS dataset with Medicare, Social Security and other administrative records and complete and correct the information regarding the death of respondents (Connor, 2006). The result of this effort is a dataset, which I will call the Center for Medicare and Medicaid Services datafile (CMS datafile), that registers the date of death or survival status for each individual that ever entered the NLTCS up to the year 2006. This dataset, however, only tracks respondents that were included in the NLTCS up to the year 1999.

The main data set I have used in all my analyses is the Duke Analytic File 2004, that includes all subjects that were ever screened for the NLTCS, along with information about all the (binary coded) ADLs (6) and IADLs (10) measured through the Community and Institutional questionnaires, for each survey year from 1982 to 2004. This file, however does not include information about age or date of births of the individuals, information that is indispensable in my modeling. I have obtained that information, together with the date of death (if applicable) from the CMS data file, which has these two pieces of information, updated until the year 2006, for individuals who *entered the sample up to the 1999 wave*. For this reason, the only information from the 2004 wave that I have included in my analyses is that from individuals who were already included in the 1999 wave.

1.2 Notation and structure

In what follows I use the following notation and structure:

1. **Subindexes**: I consider indexed variables with several subindexes. When there is no risk of ambiguity, I drop indices to refer to the whole vector, e.g. $\xi_i = (\xi_{i1}, ..., \xi_{iK})$. When I want to be more specific about the component I am dropping I replace it by an asterisk to mark its position ('*'), e.g. $\xi_{i*} = (\xi_{i1}, ..., \xi_{iK})$.
2. **Individuals**: There are N subjects in the sample, indexed by i. In the NLTCS, these individuals are the unique individuals that were registered at least once in the sample. For the whole NLTCS sample, $N \approx 49,000$ individuals.
3. **Measurements**: There are a total number T of measurement events, in which data are recorded from the individuals. I index these measurement events by the letter $t \in \{1, ..., T\}$. Each individual can be potentially measured in any of these measurement events, but is only actually measured in a subset $M_i \subset \{1, 2..., T\}$ of them. In the NLTCS the measurement events are the waves ($T = 6$) and each set M_i represent the waves at which data from that particular individual i was recorded.

4. **Responses**: For each individual $i \in \{1, ..., N\}$, at each measurement event $t \in M_i$ the sample registers J binary variables simultaneously. I index each of these J responses by the letter $j \in \{1, ..., J\}$. In the NLTCS these binary variable are the (binary) answers to the ADLs and IADLs. The realized measurement for individual i and question j at measurement time t will be denoted by $y_{ijt} \in \{0, 1\}$.

5. **Covariates**: Each individual has an associate covariate vector $X_i = (X^t_{i*}, X^f_i)$ that can be further partitioned into a time dependent component, $X^t_{i*} = (X^t_{i1}, ..., X^t_{iT})$, and a fixed component, X^f_i. An example of a time dependent component is the age at the time of measurement $t \in \{1..T\}$, Age_{it}. An example of a fixed covariate is the date of birth of the individual, DOB_i. The notation $X_{it} = (X^t_{it}, X^f)$ refers to the vector that combines the fixed part and the time dependent part at measurement time t.

6. **Sets of indexes**: I frequently need to select sets of indexes to iterate over indexed variables or other entities. To denote these sets I use the letter $\Xi^{(\cdot)}_{\cdot}$, where the subindexes are the indexes that were used to determine the grouping and the superindexes, the free indexes. For instance $\Xi^{(jt)}_{ij} = \{(i, j) : \text{some condition involving } j \text{ and } t\}$. In some cases, when no ambiguity arises, I drop these sub- and super-indexes.

To simplify notation, I use the symbol $p(\cdot)$ to denote either the probability density function or the probability mass function of the argument, as needed.

As an example, Table 1.2 shows an schematic representation of the information corresponding to a single (fictional) individual.

		Wave (t)					
		1	2	3	4	5	6
	Year	1982	1984	1989	1994	1999	2004
ADL (j)	EAT	0	0	0	0	1	1
	DRS	0	1	0	0	0	1
	TLT	0	0	0	1	1	1
	BED	1	1	0	1	1	1
	MOB	0	0	0	0	1	1
	BTH	0	0	0	0	1	1
Other	Age:	66	69	74	79	84	89
	DOB:	1916					
	DOD:	2005					
	Sex:	Female					

Table 1.2: Example of data structure for a single fictional individual. The individual itself is indexed by the letter $i \in \{1...N\}$

1.3 Modeling Functional Disability in the NLTCS: An Overview

The NLTCS is a rich source of information for questions about aging and disability in the United States and as such has been analyzed from different perspectives a number of times.

From a modeling perspective, the NLTCS functional disability data poses serious challenges to data analysts. First, responses to individual ADL and IADL disability questions are binary categorical variables and the resulting response patterns form large and extremely sparse contingency tables (e.g. the subset of 6 ADLs and 10 IADL analyzed by Erosheva et al. (2007) forms a 16 dimensional cross classification contingecy table, with $2^{16} = 65,536$ unique pattern responses, most of them with zero counts). If we consider the whole vector of responses with its six waves, the dimensionality can grow to $2^{6\times 6}$ unique patterns, just considering the ADLs, and 2^{96} including IADLs. Second, if we want to estimate changes in time, we must find a way to incorporate the longitudinal nature of the data into the analysis in a way that allows meaningful interpretation. Third, the NLTCS has a complex design and it is not clear how to incorporate the survey design and sampling weights into complex longitudinal analyses.

Manton and collaborators at the Center for Demographic Studies at Duke University used the NLTCS data to estimate the progression of the prevalence of disabilities in the United States, using successive NLTCS waves (Manton et al., 1997, 2006; Manton, 2008). They found a persistent decline of the disability from wave to wave. They compared uncorrelated cross sectional estimates of prevalence of ADLs and IADLs, using survey weights to adjust for the unequal sampling probabilities. Although they explicitly discuss the importance of the NLTCS as a source of longitudinal data they do not use the information involving the same individuals over time in their analyses.

Many of the efforts for the analysis of the functional disability data from the NLTCS have been directed to devise effective and meaningful ways of performing dimension reduction into the space of possible disability response patterns using a latent structure approach. In the literature we can distinguish two basic approaches using the latent structure idea. The first one is based on the idea of finding lower dimensional latent representations of underlying disability *states* for which the different patterns of response to ADLs and AIDLs can be taken as observable manifestations.

Thus far, the preferred method for this first approach has been the Grade of Membership (GoM) model (Woodbury et al., 1978; Manton et al., 1994; Erosheva, 2002; Erosheva et al., 2007). In the GoM model, dimensionality reduction is achieved assuming the existence of ideal pure types or "extreme profiles" and further assuming that the multivariate response of each individual can be modeled as convex combinations, with individual weights, of those extreme profiles. The extreme profiles are interpreted as reference types of disability states, where each individual could be situated *at a given time*. In this way, extreme profiles represent states whose convex combinations any person could potentially transit through the course of their life. This perspective—although without considering the time dimension—was first introduced into the analysis of disability by Manton and his collaborators at Duke University, see e.g., Manton et al. (1994). Erosheva (2002) and Erosheva et al. (2007) formulate the model in a Bayesian framework using an equivalent constrained latent class representation for the model and applied it to NLTCS pooled data from different waves. All these analyses were based on pooled cross-sectional data, without taking into account the full longitudinal nature of the actual NLTCS data.

One use of the membership scores, mainly employed by Manton and collaborators (e.g. Manton et al. 1994, 2004, 1992, 2008; Stallard 2005), treats them as a set of covariates, expressing the health status of an individual, with different purposes. In this way one can take advantage of a low-dimensional representation in order to analyze the relationship between disability and other phenomena.

White (2008) and Stallard (2005) extend this strategy of latent modeling (i.e. dimensionality reduction in the space of ADLs and IADLs response patterns) to the longitudinal setting. White (2008) formulates extensions to the latent class transition model defining a number of exclusive latent classes, representing states of disability, and estimating transition rules governing how individuals transit across classes from one measurement to the next. In this formulation, each individual at a given time is supposed to belong exclusively to a single class. White (2008) extends this idea to accommodate the NLTCS rolling enrollment. His approach deals with this issue essentially by stratifying the sample according to enrollment characteristics.

Stallard (2005) used a similar idea, but defined a set of latent non-exclusive ideal classes (in a GoM style) and transition rules depending on age (not wave), to model the transit over time of individuals within the latent unit simplex of group memberships. Stallard's approach uses Markov transition matrices, effectively treating GoM membership scores as probabilities, although interpreting them in the GoM sense.

A second approach to latent structure modeling incorporates the heterogeneity of the population by assuming the existence of different classes of *individuals*, using the latent structure to differentiate them. In this way, the latent structure does not model states of disability, as in the previous approaches, but types of people. This was the approach taken by Connor (2006), who defined latent classes of people and within each class, models the probability of acquiring a disability as a continuous monotonic function of age. In this approach, each individual is assumed to completely belong to one and only one latent class, permanently, and the analyst's task is then to characterize each class and to identify membership probabilities for the individuals.

In the next subsections I will present the Grade of Membership and the Multivariate Latent Trajectory models in more detail.

1.3.1 Mixed membership models. The Grade of Membership Model

Grade of Membership models (Woodbury et al., 1978) are well suited for large sparse contingency tables. In broad terms, they work by assuming a latent classification of the outcomes based on the idea of "partial membership": if we assume the existence of a small number of "typical cases" or "extreme profiles", then we can use them to classify the individuals in the sample according to their *relative distance* to the extremes. We allow individuals to share characteristics from each of the extremes simultaneously in different degrees, according to their position with respect to the extremes. Thus we say that the individuals "partially belong" to the profiles.

Besides applications to the study of disability (Manton et al., 1991, 1993, 1994, 2004, 2008; Stallard, 2005; Erosheva, 2002; Erosheva et al., 2007), GoM and related models have been used in applications such as network analysis (Airoldi et al., 2008), electoral preferences analysis (Gormley, 2006), estimation of judgement accuracy (Cooil and Varki, 2003), estimation of population sizes (Manrique-Vallier and Fienberg, 2008), and text classification (Erosheva et al., 2004; Blei et al., 2003; Blei and Lafferty, 2007).

To construct the GoM model we start by assuming the existence of a specific number, K, of "extreme classes" or "pure types". Suppose that there are N individuals on the population. For the ith individual, for $i \in \{1, \ldots, N\}$, we associate a J-dimensional binary vector of manifest variables $y_{i*} = (y_{i1}, ...y_{iJ})$. For any individual that is a *full member* of the kth extreme class (i.e. an "ideal" individual of the kth class), we assume that the probability of a positive response in the jth entry of the manifest variables vector is the same, i.e., $\Pr(Y_{ij} = 1 | i\text{th individual in } k\text{th class}) = \lambda_{jk}$.

We associate each individual with its own K-dimensional "membership vector", $g_{i*} = (g_{i1}, g_{i2}, ..., g_{iK})$, representing how much of a member of each class this particular individual. Membership scores are restricted so that $g_{ik} > 0$ and $\sum_{k=1}^{K} g_{ik} = 1$. We will call this geometric place the "$K - 1$ dimensional unit simplex" and denote it by the symbol Δ_{K-1} or Δ, if no ambiguity arises. We introduce the idea of "partial membership" by setting the distribution of each manifest variable given the membership vector as the convex combination,

$$p(y_{ij}|g_{i*}) = \sum_{k=1}^{K} g_{ik} \lambda_{jk}^{y_{ij}} \left(1 - \lambda_{jk}\right)^{1-y_{ij}}.$$

We further assume that the item responses j are conditionally independent given membership vectors. This condition, sometimes referred as *latent conditional independence* or *local independence* (Holland and Rosenbaum, 1986; Sijtsma and Junker, 2006), expresses the idea that the membership vector g completely explains the dependence structure between the J binary manifest variables. By making this assumption, we can construct the conditional joint distribution of responses

$$p(y_{i*}|g_{i*}) = \prod_{j=1}^{J} \sum_{k=1}^{K} g_{ik} \lambda_{jk}^{y_{ij}} \left(1 - \lambda_{jk}\right)^{1-y_{ij}}.$$

Further assuming that the individuals are randomly sampled from the population we finally get,

$$p(y_{**}|g_{**}) = \prod_{i=1}^{N} \prod_{j=1}^{J} \sum_{k=1}^{K} g_{ik} \lambda_{jk}^{y_{ij}} \left(1 - \lambda_{jk}\right)^{1-y_{ij}}. \tag{1.1}$$

It is worth noting that the GoM model can be understood as a generalization of a more traditional latent class model (LCM). See, Goodman (1974) for a description of these. The expression in Equation (1.1) can also be used to represent the likelihood conditional (on class membership) of an LCM by restricting the vector g_{i*} so that $g_{ik} = 1$ for some $k \in \{1...K\}$ and $g_{ik'} = 0$ for every $k' \neq k$ (Erosheva, 2005). This is, every individual is a full member of a single class.

Erosheva (2002) and Erosheva et al. (2007) formulate the GoM model in a Bayesian framework and apply it to pooled NLTCS data. Following an observation by Haberman (1995), Erosheva (2002) shows that, even though GoM models can be thought as a generalization of a LCM, GoM models themselves can be also characterized as a special subclass of some restricted LCM models. This somewhat surprising observation provided the basis for the estimation methods in Erosheva (2002) and Erosheva et al. (2007).

Consider the membership scores g_i as i.i.d. samples from a common population-level distribution, $g_i \overset{iid}{\sim} G$. Under that assumption,

$$\begin{aligned} p(y_i) &= \int_{\Delta} \prod_{j=1}^{J} \sum_{k=1}^{K} g_k \lambda_{jk}^{y_{ij}} (1-\lambda_{jk})^{1-y_{ij}} G(dg) \\ &= \sum_{z \in Z} \pi_z \lambda_{jk}^{y_{ij}} (1-\lambda_{jk})^{1-y_{ij}}, \end{aligned} \tag{1.2}$$

where

$$\begin{aligned} Z =& \{1, 2, ..., K\}^J \\ \pi_z =& E_G \left[\prod_{j=1}^{J} \prod_{k=1}^{K} g_k^{I(z_j=k)} \right] \end{aligned} \tag{1.3}$$

The expression in Equation (1.2) is typical of a discrete mixture model. What makes the GoM something other than just a LCM with a potentially huge number of components ($\#Z = K^J$) is that (a) π_z is invariant to permutations of the components of z, giving an effective number of just $(K + J - 1)!/(J!(K - 1)!) - 1$ different values of π to estimate, and (b) the π values are constrained to be the expected values given by Equation (1.3) (Haberman, 1995).

In the rest of this thesis I will make widespread use of this representation.

All the applications of the GoM to the NLTCS have described the extreme profiles as (extreme) “states of disability” and the membership scores as the way each individual *at a given time* is situated within the convex hull defined by the extreme profiles (Erosheva et al., 2002). This defines the particular disability status of that individual.

1.3.2 Modeling Longitudinal Patterns - Latent Trajectory Models

Latent trajectory models (Nagin, 1999) have become increasingly popular in criminology and other social sciences as a method to analyze heterogeneous longitudinal data where homogeneous sub populations of individuals are suspected to exist (see e.g. Kreuter and Muthén 2008; Nagin and Land 1993).

Different from latent transition models (e.g. White 2008; Stallard 2005), trajectory models seek to represent the whole progression (or trajectory) of the response over time. To this end, we introduce the concept of “trajectories”: parametric functions dependent on time that describe the evolution of the phenomenon under consideration. Connor (2006) extends the latent trajectory class of models to binary multivariate longitudinal data to analyze the NLTCS.

The construction of Connor’s Multivariate Latent Trajectory model (MVLT model) is similar to the specification of other Latent Class or discrete mixture models. Assume that the population can be partitioned into K unobserved exclusive classes. Then, assume that individuals that belong to those classes can be considered homogeneous with respect to the characteristics to be modeled. In the present instance, those characteristics are the trajectories of the probability of a positive response on each of the response variables. Connor (2006) uses a logit link function to model the frailty trajectory within each class as a monotone function of age. Call $z_{i*} = (z_{i1}, ..., z_{iK})$ the indicator vector for the individual indexed by i. An indicator vector for individual i is a K-dimensional vector $z_{i*} = (z_{i1}, ... z_{iK})$ such that $z_{ik} = 1$ and $z_{ik'} = 0$ for $k' \neq k$ if and only if the individual indexed by i belongs to class $k \in \{1, 2, ..., K\}$. Then, specify the trajectory of the probability of acquiring a disability in ADL-j, for members of the class k as a function of Age,

$$\begin{aligned} \lambda_{jk}(Age_{it}) &= \Pr(Y_{ijt} = 1 | z_{ik} = 1, Age_{it}) \\ &= \text{logit}^{-1}(\beta_{0jk} + \beta_{1jk} \cdot Age_{it}) \end{aligned} \tag{1.4}$$

so that

$$\begin{aligned} \Pr(Y_{ijt} = y_{ijt} | z_{ik} = 1, Age_{it}) &= f_{\theta_j}(y_{ijt} | Age_{it}, z_i) \\ &= \lambda_{jk}(Age_{it})^{y_{ijt}} \left(1 - \lambda_{jk}(Age_{it})\right)^{1 - y_{ijt}} \end{aligned}$$

Then, assuming that the probability of *any* individual of being a member of class k is π_k, i.e. $\Pr(z_{ik} = 1) = \pi_k$, and assuming conditional independence between the responses to different questions and measurement events given class membership, we can construct the joint distribution of the responses as

$$\Pr(Y_i = y_i | Age_i) = \sum_{k=1}^{K} \pi_k \prod_{j=1}^{J} \prod_{t=1}^{T} f_{\theta_{jk}}(y_{ijt} | Age_{it}).$$

Finally, assuming random sampling, the joint probability model becomes,

$$\Pr(Y = y | Age) = \prod_{i=1}^{N} \sum_{k=1}^{K} \pi_k \prod_{j=1}^{J} \prod_{t=1}^{T} f_{\theta_{jk}}(y_{ijt} | Age_{it}) \tag{1.5}$$

Connor (2006) applies these models to the NLTCS data in order to estimate the trajectories of the probability of acquiring disabilities as a function of personal time or age. The latent clustering framework allows some form of heterogeneity modeling, assuming that there are exactly K different trajectories that each individual could in principle follow, and estimating the proportion of the population that belongs to each of these classes, π_k.

1.4 Overview of the thesis

This thesis is organized around the problem of modeling the NLTCS data by combining the ideas of longitudinal trajectories and mixed membership classification. Chapter 2 introduces and develops what I call the Basic Trajectory Grade of Membership model, a model that describes typical trajectories over time and handles heterogeneity of the population using a GoM-like approach. I specialize the model to handle the NLTCS data and, in Chapter 3 I apply it to the modeling of disability.

The next four chapters describe extensions of the Basic model aimed to represent more subtle aspects of the NLTCS, and report on their application to the real data. Chapter 4 extends the Basic model to detect changes in the ways of aging across birth cohorts. Chapter 6 introduces an extension that seeks to simultaneously model individuals' disability and survival outcomes, to investigate the interplay between disability and survival and to improve the general quality of the estimates, allowing the use of disability and survival data to estimate a common pool of parameters. Chapter 5 and 7 report on the application of these two extensions to the NLTCS data.

Chapter 8 recapitulates the contributions made by this thesis and proposes a program for future research.

Chapter 2

Basic Grade of Membership Trajectory Model

In this chapter I introduce a new longitudinal model which aims to characterize the joint evolution of the probabilities of a set of binary responses as a function of time in presence of individual-level heterogeneity. In the application to the study of disability using the NLTCS data, my aim is to model the evolution of the probability of acquiring specific disabilities as a function of personal time (time in the system or age).

The modeling approach I take in this chapter is based on the concept of Latent Trajectory (see Connor 2006). I model the evolution of the underlying probability of a positive response to each binary variable by determining underlying functions of time that govern those probabilities. At the same time I handle the heterogeneity present in the population in a flexible and interpretable way.

As several authors have noted, latent trajectory models, such as the ones proposed in Nagin (1999) and Connor (2006), assume the existence of latent homogeneous populations to which potentially no individuals actually belong (see e.g., Kreuter and Muthén (2008)). Even under a non-committed interpretation, understanding these models more as an approximation device to the true underlying distribution, than as the discovery of "true" latent populations, we still have the problem of attributing all the variability on the actual individual trajectories to random fluctuations within a class, disregarding the fact that each class might be just too broad a description for each individual.

One possible solution to this situation, within the Latent Trajectory framework, is to increase the number of latent classes, making enough room for more variability attributable to true underlying differences across individuals. This solution, however, has the problem of making the models more complex, sometimes by introducing new and only slightly different classes to account for observable differences.

My proposed solution takes a different approach. I handle heterogeneity by using the Grade of Membership model ideas, considering each multivariate *trajectory* as a simultaneous combination of a number, K, of "extreme profiles" or "pure types". In this way, similar to Connor (2006), these models construct a clustering of individuals within the population of interest and describe their prototypical multivariate trajectories over time, but borrowing the GoM ideas, they regard these trajectories as ideal cases and assume a *soft clustering* over the population of interest, where each individual is allowed to belong to more than one extreme profiles simultaneously, in different degrees.

It is important to distinguish this approach from other longitudinal applications of latent class or GoM models. In other applications to the study of disability, such as those in

Stallard (2005) or White (2008), the (soft) group membership involves states (or "fuzzy states" in the GoM variants) that change over time with the intention of estimate transition from one point in time to another. In this setup I am regarding the group membership of each individual as fixed over time. In this sense, my approach attempts to characterize the population by describing ideal reference types of *people*.

2.1 General Formulation

Similar to the construction of the GoM model (see Section 1.3.1), I start assuming the existence of a small number, K, of extreme profiles or ideal types of individuals and, for these ideal types, model the evolution of the marginal distribution of each response variable, y_{ij*}, as a function of some individual time-dependent covariates, X_{it}. We will call the expected values of these functions "extreme trajectories", $\lambda_{jk}(X_{it})$.

Each real (as opposed to ideal) individual will be considered as a unique mixture of the K extreme profiles, using individual weights arranged on a "membership vector", $g_{i*} = (g_{i1}..., g_{iK})$ ($g_{ik} > 0$, $\sum_k g_{ik} = 1$). I will consider the membership vector as a characteristic of the individual that is fixed for each individual, but unknown. The interpretation for these weights, g_{ik}, is that they represent the "degree of membership" of individual i in class K. Ideal individuals will be characterized by a membership vector where the component corresponding to the profile where they (exclusively) belong has a value $g_{ik} = 1$, and the rest of entries, zeroes. For instance, an individual with membership vector $g_i = (0, 1, 0, 0)$ belongs exclusively to the extreme profile $k = 2$.

In this way, the distribution for response component j, for extreme profile k can be written as:

$$\Pr\left(Y_{ijt} = y_{ijt} | G_{ik} = 1, X_i\right) = f_{\theta_{j|k}}\left(y_{ijt} \mid X_{it}\right)$$

so that

$$\lambda_{jk}(x) = E[f_{\theta_{j|k}}\left(Y_{ijt} \mid x\right)].$$

Using the membership vectors, we will model the corresponding trajectory for a generic, non-ideal individual i, with membership vector $G_{i*} = (g_{i1}, ..., g_{iK})$ as the convex combination,

$$\Pr\left(Y_{ijt} = y_{ijt} \left| G_{i*} = (g_{i1}, ..., g_{iK}), X_{i*} \right.\right) \;=\; \sum_{k=1}^{K} g_{ik} f_{\theta_{j|k}}(y_{ijt}|X_{it}) \tag{2.1}$$

Next, we will introduce a local independence assumption: for a single individual, conditional on the covariate X_i and its membership vector, g_{i*}, the J responses at each of the T measurement times are independent of each other. By this assumption we are effectively using the membership vector and the covariates to decouple the dependence structure present in the components of the response:

$$\Pr\left(Y_{i**} = y_{i**} \left| G_{i*} = (g_{i1}, ..., g_{iK}), X_{it} \right.\right) \;=\; \prod_{j=1}^{J} \prod_{t=1}^{T} \sum_{k=1}^{K} g_{ik} f_{\theta_{j|k}}(y_{ijt}|X_{it}) \tag{2.2}$$

By assuming that each individual has been randomly sampled from the population we finally get the joint model, conditional on g and X,

$$\Pr\left(Y_{***} = y_{***} \left| g_{**}, X_{**} \right.\right) \;=\; \prod_{i=1}^{N} \prod_{j=1}^{J} \prod_{t=1}^{T} \sum_{k=1}^{K} g_{ik} f_{\theta_{j|k}}(y_{ijl}|X_{it}) \tag{2.3}$$

If we assume the membership vectors are i.i.d. samples from a common distribution with support in Δ_{K-1}, say G_α, we can finally get the unconditional model,

$$\Pr\left(Y = y|X\right) \quad = \quad \prod_{i=1}^{N} \int_{\Delta} \prod_{j=1}^{J} \prod_{t=1}^{T} \sum_{k=1}^{K} g_k f_{\theta_{j|k}}(y_{ijl}|X_{it}) G(dg) \tag{2.4}$$

We can gain further insight on this formulation through the application of the Haberman-Erosheva constrained LCM equivalence (see Section 1.3.1). Applying it we get

$$\Pr\left(Y = y|X\right) \quad = \quad \prod_{i=1}^{N} \sum_{z \in Z} \pi_z \prod_{j=1}^{J} \prod_{t=1}^{T} f_{\theta_{j|k}}(y_{ijl}|X_{it})$$

where

$$\begin{aligned} Z &= \{1, 2, ..., K\}^{J \times T} \\ \pi_z &= E_G\left[\prod_{j=1}^{J} \prod_{t=1}^{T} \prod_{k=1}^{K} g_k^{I(z_{jt}=k)}\right] \qquad \text{for } z \in Z \end{aligned} \tag{2.5}$$

This representation shows a direct connection with the formulation in Connor (2006), (see Section 1.3.2). The biggest difference between these two formulations is that instead of having K unconstrained profile weights as in Connor (2006), we have $K^{J \times T}$ of them, strongly constrained by the moments formulation implicit in Equation (2.5) (see Section 1.3.1 for further details). This fact will allow us model more complex data structures with just a few extreme profiles, at the price of having to estimate the partial membership structure.

Our formulation also suggests that we can perform inferences based on the augmented-data likelihood

$$p\left(y, z \,|, X, \theta, g\right) \;=\; \prod_{i=1}^{N}\prod_{j=1}^{J}\prod_{t=1}^{T}\prod_{k=1}^{K}\Big[g_{ik} f_{\theta_{j|k}}(y_{ijl}|X_{it})\Big]^{I(z_{ijt}=k)} \qquad (2.6)$$

through the use of a variants of the data augmentation algorithm proposed in Erosheva (2002) and Erosheva et al. (2007).

2.1.1 What To Do With The Missing Values?

The generic likelihood of the Basic model in Equation (2.6) is assuming explicitly that the vector y_{i**} is rectangular, with j ranging from 1 to J, and t from 1 to T for every single individual $i = 1...N$. In reality however, for some individuals one or many of the measurements are not present. Reasons for this include attrition, either due to death or other reasons, being ineligible in some of the first waves, but not in later ones, according to the sample design; and late entry, in rolling enrollment panel designs.

In the case of the NLTCS, ineligibility in some early waves is a major source of lack of rectangularity of the response: individuals have to be older than 65 to qualify and that requirement might not be met until late waves, when they are incorporated to make up for those who have died. The second and most important cause of missing measurements is death (Clark, 1998; White, 2008). Distinguishing death from other causes of missingness can be critical, since death is very likely to be highly correlated with the outcome of interest.

While there are individuals who have missing measurements in the middle of the measurement sequence, for instance, entered in wave 2, skipped wave 3, and have continuous measurements until wave 6, this occurrence is quite infrequent (White, 2008).

For the reminder of this chapter I rely on the strong assumption that the missing measurements are *Missing at Random* (MAR) (Rubin, 1976). In essence, the MAR hypothesis states that the reason why a measurement is missing is not dependent on the outcome that we are trying to model nor on the parameters of that model. Under these conditions, inferences based on posterior distributions given the observed outcomes, completely ignoring the missing data, will lead to the same conclusions that would have been obtained had we explicitly modeled the missing data and averaged it out (Rubin, 1976; Gelman et al., 2004).

While the MAR assumption might be reasonable for the missingness due to non-eligibility, given that we are already conditioning on its primary cause, the age of the subjects, it is highly unlikely that this is the case for the missingness due to the subjects' death. We will return to this point and analyze the consequences of this assumption by addressing it explicitly in Chapter 6.

Under the MAR hypothesis, the estimates that we would have obtained by modeling the missing data mechanism (essentially, averaging the non-observed values according to the model and conditioning on the data and observed pattern of missingness), will be equivalent to those obtained by replacing the augmented data likelihood in Equation (2.6) by

$$p\left(y, z\,|, X, \theta, g\right) \;=\; \prod_{i=1}^{N}\prod_{j=1}^{J}\prod_{t\in M_i}\prod_{k=1}^{K}\Big[g_{ik} f_{\theta_{j|k}}(y_{ijl}|X_{it})\Big]^{I(z_{ijt}=k)} \tag{2.7}$$

This is, by replacing the individual product over all waves, $\prod_{t=1}^{T}\cdot$, by the product over all the waves where the individual was present, $\prod_{t \in M_i} \cdot$.

2.2 Continuous Trajectories GoM Model

The choice of trajectory functions $\lambda_{jk}(\cdot)$ must be application-specific, as it will encode assumptions about the nature of the underlying process, expressed as the realizations of the response variables. For the application to the NLTCS, following Connor (2006), I will use the s-shaped monotonic functions

$$\begin{aligned} \lambda_{j|k}(X_{it}) =& \frac{\exp(\beta_{0jk} + \beta_{1jk} Age_{it})}{1 + \exp(\beta_{0jk} + \beta_{1jk} Age_{it})} && (2.8) \\ =& \text{logit}^{-1} \left(\beta_{0jk} + \beta_{1jk} Age_{it} \right) && (2.9) \end{aligned}$$

which will lead to the conditional distribution

$$\begin{aligned} f_{\theta_{j|k}}(y_{ijt}|X_{i*}) &= \text{Bernoulli}\left(y_{ijt}|\lambda_{j|k}(X_{it})\right) && (2.10) \\ &= \lambda_{j|k}(Age_{it})^{y_{ijt}} \left(1 - \lambda_{j|k}(Age_{it})\right)^{1-y_{ijt}}, && (2.11) \end{aligned}$$

where Age_{it} is the age of the ith individual at measurement time t. This specification has the advantage of being relatively simple, with just $2 \times J$ parameters per extreme profile and of representing the intuitively sound notion that the underlying probability of disability is a monotonic (increasing) function of age. Other specifications are certainly possible and I will present an alternative one in Section 2.3 of this Chapter.

Similar to Erosheva (2002) and Airoldi et al. (2008), I take the common distribution of the N membership vectors g_i, G_α, as

$$g_{i*}|\alpha \overset{iid}{\sim} \text{Dirichlet}(\alpha).$$

where $\alpha = (\alpha_1, \alpha_2, ..., \alpha_K)$ with $\alpha_k > 0$ for all $k = 1, ..., K$ and

$$p(g|\alpha) = \frac{\Gamma\left(\sum_{k=1}^{K} \alpha_k\right)}{\prod_{k=1}^{K} \Gamma(\alpha_k)} \prod_{k=1}^{K} g_k^{\alpha_k - 1} \cdot I(g \in \Delta_{K-1})$$

The Dirichlet distribution in this setting has some good properties. In the first place, it is conjugate to the multinomial distribution, facilitating a great deal the computations using Gibbs samplers; second, adopting the re-parametrization $\alpha = (\alpha_0 \cdot \xi_1, ..., \alpha_0 \cdot \xi_K)$ with $\alpha_0 > 0$, $\xi_k > 0$ and $\sum_k \xi_k = 1$ we can interpret the vector ξ as the average proportion of the population in the k-th extreme profile and, α_0 as a parameter governing the spread of the distribution: as α_0 approaches 0, the samples from G_α are more and more concentrated on the vertices of Δ_{K-1} and; as α_0 increases they are more concentrated near the average, ξ.

As Erosheva et al. (2007) and Airoldi et al. (2007) discuss, a priori setting the parameters α for the Dirichlet distribution might be too strong an assumption to do realistic modeling. I will try to estimate these parameters specifying hyper-priors and computing posterior distributions. For this purpose I use hyper priors for α_0 and ξ similar to the ones in Erosheva (2002) and Erosheva et al. (2007):

$$\begin{aligned} \alpha_0 &\sim \text{Gamma}(\tau, \eta) \\ \xi &\sim \text{Dirichlet}(\mathbf{1}_K) \quad \text{(Uniform over } \Delta_{K-1}). \end{aligned}$$

This specification takes advantage of the interpretation of the parameters α_0 and ξ, by considering them independent characteristics that we can model separately. For the same reason I will assume that $p(\alpha_0, \xi) = p(\alpha)p(\xi)$.

For the parameters that characterize the trajectories, β_0 and β_1 I have chose the priors

$$\beta_{0jk} \overset{iid}{\sim} N(\mu_0, \sigma_0^2) \qquad \text{and} \qquad \beta_{1jk} \overset{iid}{\sim} N(\mu_1, \sigma_1^2),$$

for all $j = 1, ..., J$ and $k = 1, ..., K$, with β_0 independent from β_1. This form of priors can in principle be set to be noninformative, by apriori specifying high variances. I have also assumed β apriori independent of α.

2.2.1 Estimation via Markov Chain Monte Carlo

Under the specification of extreme trajectories and priors outlined in this section, and following the augmented data representation in Equation (2.7), the joint posterior distribution of parameters and augmented data is

$$\begin{aligned} p(\alpha, \beta, g|y, z, Age) \propto & p(\beta)p(\alpha) \left(\prod_{i=1}^{N} p\left(g_{i*}|\alpha\right) \right) \\ & \times \prod_{i=1}^{N} \prod_{j=1}^{J} \prod_{t \in M_i} \prod_{k=1}^{K} \left(g_{ik} \frac{\exp(y_{ijt}\beta_{0jk} + y_{ijt}\beta_{1jk}Age_{it})}{1 + \exp(\beta_{0jk} + \beta_{1jk}Age_{it})} \right)^{I(z_{ijt}=k)}, \end{aligned}$$

with $p(\alpha) = p(\alpha_0) \cdot p(\xi)$ and

$$\begin{aligned} p(g_{i*}|\alpha) &= \text{Dirichlet}(g_{i*}|\alpha_1, \alpha_2, ..., \alpha_k), \\ p(\alpha_0) &= \text{Gamma}(\alpha_0|\tau, \eta), \\ p(\xi) &= \text{Dirichlet}\big(\xi|\mathbf{1}_K\big) \quad \text{(Uniform on the } \Delta_{K-1}), \end{aligned}$$

and $\alpha_0 = \sum_k \alpha_k$ and $\xi = (\xi_1, \xi_2, ..., \xi_K)$ with $\alpha_k = \alpha_0 \cdot \xi_k$. Parameters τ and η are shape and inverse scale parameters, respectively.

Specifying an MCMC algorithm to obtain approximate realizations from this posterior using the Gibbs sampling algorithm (Geman and Geman, 1984; Casella and George, 1992) is just matter of obtaining the full conditional distributions for each parameter and augmented data (for data augmentation algorithm see Tanner 1996; Marin and Robert 2007; Robert and Casella 2004). Unfortunately, many of the resulting full conditional distributions have non standard forms and therefore more sophisticated approaches are needed. The basic structure that we will use is the block-metropolis (Chib, 2004), "Metropolis-Hastings within Gibbs" (Gamerman, 1997) or "Hybrid Gibbs-Metropolis" (Robert and Casella, 2004). This algorithm consists in dividing the multivariate parameter vector into fixed multivariate or univariate blocks and proceed as in the case of the Gibbs algorithm: sequentially sampling from the joint distribution of each of these blocks conditional on the current values of the other parameters. If any of those blocks, say x, happens to have a non-standard distribution, we insert a *Metropolis-Hastings step*:

1. Conditional on the current values of *all* variables, obtain a proposal value x^* from a fixed distribution that is allowed to depend on the current value of all parameters, including x, $q(x^*|x, ...)$.

2. Compute
$$r = \min\left\{ \underbrace{\frac{p(x^*|...)}{p(x|...)}}_{r_M} \times \underbrace{\frac{q(x|x^*,...)}{q(x^*|x,...)}}_{r_H}, 1 \right\}.$$
   The second part of the ratio, r_H, compensates for the asymmetry of the proposal distribution. When the proposal distribution is symmetric (i.e. $q(x|x^*,...) = q(x^*|x,...)$), it cancels out leaving only the r_M component. In that case we call this a "Metropolis step".

3. Update the chain, from step m to step $m+1$ according to :
$$x^{(m+1)} = \begin{cases} x* & \text{with probability } r \\ x^{(m)} & \text{with probability } 1-r \end{cases}$$

This is the standard Metropolis-Hastings algorithm, but where we are just applying it into a single step for sampling from the difficult full conditional distribution $p(x|...)$. The important thing to note is that the proposal and rejection step is just performed once per updating. Standard results (Gamerman, 1997; Marin and Robert, 2007; Robert and Casella, 2004; Robert, 2004) guarantee that this chain converges to the desired stationary distribution. While certainly other more sophisticated approaches are possible, and I have indeed tried a number of them, like the adaptive rejection Metropolis-Hastings (Gilks and Best, 1992) or the multiple-try Metropolis-Hastings (Liu, 1999), the simple application of MH steps described below outperforms these variants because of their relative low computational cost.

Now we can describe the algorithm step by step.

1. **Sampling from z (imputation)** For every $i \in \{1 \dots N\}, j \in \{1 \dots J\}$ and $t \in M_i$,

sample

$$z_{ijt}|... \sim \text{Discrete}(p_1, p_2, ..., p_K)$$

with

$$p_k \propto g_{ik} \frac{\exp(y_{ijt}\beta_{0jk} + y_{ijt}\beta_{1jk}Age_{it})}{1 + \exp(\beta_{0jk} + \beta_{1jk}Age_{it})}$$

for all $k \in \{1, \ldots, K\}$.

2. **Sampling from** $(\beta_{0jk}, \beta_{1jk})$: The full joint conditional distribution of $(\beta_{0jk}, \beta_{1jk})$ is

$$\begin{aligned}
p\left(\beta_{0jk}, \beta_{1jk}|...\right) &\propto p(\beta_{0jk}, \beta_{1jk}) \times \prod_{i,t} f(y_{ijt}|\beta_{0jk|}, \beta_{1jk}, Age_{it})^{I(z_{ijt}=k)} \\
&\propto p(\beta_{0jk}, \beta_{1jk}) \times \prod_{i,j} \left[\frac{\exp\left[\left(\beta_{0jk} + \beta_{0jk}Age_{it}\right) y_{ijt}\right]}{1 + \exp\left(\beta_{0jk} + \beta_{0jk}Age_{it}\right)}\right]^{I(z_{ijt}=t)} \\
&\propto \left[\prod_{jt} \left[1 + \exp\left(\beta_{0jk} + \beta_{0jk}Age_{it}\right)\right]^{I(z_{ijt}=k)}\right]^{-1} \\
&\quad \times \exp\Bigg[-\frac{{\beta_{0jk}}^2}{2{\sigma_0}^2} + \beta_{0jk}\left(\frac{\mu_0}{\sigma_0} + \sum_{ij} I(z_{ijt} = k)y_{ijt}\right) \\
&\qquad - \frac{{\beta_{1jk}}^2}{2{\sigma_1}^2} + \beta_{1jk}\left(\frac{\mu_1}{\sigma_1} + \sum_{ij} I(z_{ijt} = k)Age_{it}y_{ijt}\right)\Bigg].
\end{aligned}$$

This expression can be written in a more compact form if we define the set of indexes

$\Xi = \Xi_{jk}^{(it)} = \{(i,t) : z_{ijt} = k\}$. Then the previous expression becomes

$$
\begin{aligned}
&p\left(\beta_{0jk}, \beta_{1jk} | ...\right) \\
&\quad \propto \frac{\exp\left[-\frac{{\beta_{0jk}}^2}{2{\sigma_0}^2} + \beta_{0jk}\left(\frac{\mu_0}{\sigma_0} + \sum_{\Xi} y_{ijt}\right) - \frac{{\beta_{1jk}}^2}{2{\sigma_1}^2} + \beta_{1jk}\left(\frac{\mu_1}{\sigma_1} + \sum_{\Xi} Age_{it} y_{ijt}\right)\right]}{\prod\limits_{\Xi}\left[1 + \exp\left(\beta_{0jk} + \beta_{0jk} Age_{it}\right)\right]}
\end{aligned}
$$

This distribution does not have a recognizable form. Thus we use a random walk Metropolis step:

(a) Proposal step: Sample the proposal values

$$\beta_{0jk}^* \sim N(\beta_{0jk}, \sigma_{\beta 0}^2) \text{ and } \beta_{1jk}^* \sim N(\beta_{1jk}, \sigma_{\beta 1}^2)$$

where the values $\sigma_{\beta 0}^2$ and $\sigma_{\beta 1}^2$ are tuning parameters that we have to calibrate to achieve a good balance between acceptance of proposed values and exploration of the support of the target distribution (Robert and Casella, 2004). Note that this proposal is symmetric, so we just need to add a Metropolis step.

(b) Acceptance step: compute

$$
\begin{aligned}
r_M &= \frac{p(\beta_{0jk}^*, \beta_{1jk}^* | ...)}{p(\beta_{0jk}, \beta_{1jk} | ...)} \\
&= \prod_{\Xi}\left[\frac{1 + \exp\left[\beta_{0jk} + \beta_{0jk} Age_{it}\right]}{1 + \exp\left[\beta_{0jk}^* + \beta_{0jk}^* Age_{it}\right]}\right] \\
&\quad \times \exp\left[-\frac{\beta_{0jk}^{*2} - \beta_{0jk}^2}{2\sigma_0^2} + \left(\beta_{0jk}^* - \beta_{0jk}\right)\left(\frac{\mu_0}{\sigma_0} + \sum\nolimits_{\Xi} y_{ijt}\right)\right] \\
&\quad \times \exp\left[-\frac{\beta_{1jk}^{*2} - \beta_{1jk}^2}{2\sigma_1^2} + \left(\beta_{1jk}^* - \beta_{1jk}\right)\left(\frac{\mu_1}{\sigma_1} + \sum\nolimits_{\Xi} y_{ijt} Age_{it}\right)\right] \qquad (2.12)
\end{aligned}
$$

and make

$$(\beta_{0jk}, \beta_{1jk})^{(m+1)} = \begin{cases} (\beta^*_{0jk}, {\beta^*}_{1jk}) & \text{with probability } \min\{r_M, 1\} \\ (\beta_{0jk}, \beta_{1jk})^{(m)} & \text{with probability 1 - } \min\{r_M, 1\}. \end{cases}$$

The reasons for blocking β_{0jk} together with β_{1jk} in a single updating step, as opposed to considering two separate updating steps, are mostly practical: in this way we only have to compute one rejection ratio and tune only one proposal parameter. This allows us to greatly optimize the number of computations needed to sample all the β parameters for all the combination of the indexes j and k (see Section 2.A.1 for details). Additionally, the posterior correlation between these two parameters can be extremely high, in the order of $|\rho| = 0.9$, for some pairs (j, k), so a well designed proposal will in principle mix the chain faster by improving the exploration of the support (Gamerman, 1997; Gilks and Roberts, 1996; Chib, 2004; Robert, 2004), which is indeed the case of this sampler. While specially designed proposal distributions, that move along areas of high joint probability would in principle work better, this algorithm works sufficiently well in practice to justify the extra complications.

3. **Sampling from g_i:** Since the Dirichlet distribution is conjugate to the multinomial, this expression is particularly simple:

$$g_i|\ldots \overset{indep.}{\sim} \text{Dirichlet}\Bigg(\alpha_1 + \sum_{j,t} I(z_{ijt} = 1), \\ \alpha_2 + \sum_{j,t} I(z_{ijt} = 2), \ldots, \alpha_K + \sum_{j,t} I(z_{ijt} = K)\Bigg).$$

4. **Sampling from α:** The full conditional distribution of α,

$$\begin{aligned}p(\alpha|...) \propto & \text{Gamma}(\alpha_0|\tau,\eta) \times \text{Dirichlet}(\xi|\mathbf{1}_K) \times \prod_{i=1}^{N} \text{Dirichlet}(g_i|\alpha) \\ \propto & \alpha_0^{\tau-1} \exp[-\alpha_0\eta] \times \prod_{i=1}^{N} \frac{\Gamma(\alpha_0)}{\prod_{k=1}^{K}\Gamma(\alpha_k)} \prod_{k=1}^{K} g_{ik}^{\alpha_k} \\ \propto & \alpha_0^{\tau-1} \exp[-\alpha_0\eta] \times \left[\frac{\Gamma(\alpha_0)}{\prod_{k=1}^{K}\Gamma(\alpha_k)}\right]^N \prod_{k=1}^{K}\left[\prod_{i=1}^{N} g_{ik}\right]^{\alpha_k}, \end{aligned} \tag{2.13}$$

does not have any recognizable form. Thus we will again use a Metropolis-Hastings within Gibbs step. Erosheva (2002) proposed a two-stage sampling strategy, obtaining samples for α_0 and for ξ through the use of two separate Metropolis-Hastings steps. I have achieved better results in terms of numerical stability, speed of convergence and computational simplicity treating the vector α as a block by using the following logarithmic scale Gaussian random walk Metropolis-Hastings step, first proposed in Manrique-Vallier and Fienberg (2008):

(a) (Proposal step) Sample $\alpha^* = (\alpha_1^*, \alpha_2^*, ..., \alpha_K^*)$, as independent lognormal variates from

$$\alpha_k^* \overset{indep.}{\sim} \text{lognormal}(\log\alpha_k, \sigma^2).$$

The lognormal(μ, σ^2) distribution is the distribution of $\exp X$, when X are is a normal variate with expected value μ and variance σ^2. It has density

$$p(x|\mu,\sigma^2) = \frac{1}{x\sqrt{2\pi\sigma^2}} \exp\left[-\frac{(\log x - \mu)^2}{2\sigma^2}\right],$$

and expected value $E[X] = \exp[\mu + \sigma^2/2]$ and variance $V[X] = (\exp\sigma^2 - 1)\exp[2\mu + \sigma^2]$. Note that although this proposal is inspired by the idea of producing a random walk in a log-transformed space, in the untransformed scale it is not centered on the current value of α. In fact for this particular proposal, we have that $E[\alpha_k^*|\alpha_k] = \alpha_k \cdot e^{\sigma^2/2}$. Different from the proposal in Step 2, this distribution is not symmetric. Thus, will have to include the Hastings correction term, r_H.

(b) (Acceptance step) Let $\alpha_0^* = \sum_{k=1}^K \alpha_k^*$. Then,

$$\begin{aligned} r_M =& \frac{p(\alpha^*|...)}{p(\alpha|...)} \\ =& \exp\left[-\eta(\alpha_0^* - \alpha_0)\right] \left(\frac{\alpha_0^*}{\alpha_0}\right)^{\tau-1} \times \left[\frac{\Gamma(\alpha_0^*)}{\Gamma(\alpha_0)} \prod_{k=1}^K \frac{\Gamma(\alpha_k)}{\Gamma(\alpha_k^*)}\right]^N \prod_{k=1}^K \left(\prod_{i=1}^N g_{ik}\right)^{\alpha_k^* - \alpha_k}, \\ r_H =& \frac{q(\alpha|\alpha^*)}{q(\alpha^*|\alpha)} = \prod_{k=1}^K \frac{\alpha_k^*}{\alpha_k}, \end{aligned}$$

and

$$\begin{aligned} r &= \min\{r_M \times r_H, 1\} \\ &= \min\Bigg\{1, \exp\left[-\tau(\alpha_0^* - \alpha_0)\right] \left(\prod_{k=1}^K \frac{\alpha_k^*}{\alpha_k}\right) \left(\frac{\alpha_0^*}{\alpha_0}\right)^{\tau-1} \\ &\qquad \times \left[\frac{\Gamma(\alpha_0^*)}{\Gamma(\alpha_0)} \prod_{k=1}^K \frac{\Gamma(\alpha_k)}{\Gamma(\alpha_k^*)}\right]^N \prod_{k=1}^K \left(\prod_{i=1}^N g_{ik}\right)^{\alpha_k^* - \alpha_k} \Bigg\}. \end{aligned}$$

Finally, update the chain, from step m to step $m+1$ according to the rule

$$\alpha^{(m+1)} = \begin{cases} \alpha^* & \text{with probability} \quad r \\ \alpha^{(m)} & \text{with probability} \quad 1 - r. \end{cases}$$

The application of this algorithm consists of successive applications of steps 1 to 4.

2.3 A Variant: Discrete Piecewise Constant Trajectories

While the sigmoid trajectory function specification in Section 2.2 seems to be both reasonable and simple to describe, it can be useful to test how the implicit constraints imposed by its strong parametric characteristics might affect the estimation.

One simple way to assess this effect is to provide another, less constrained specification in order to compare the results. In this section, I replace the specification of the extreme trajectories, $\lambda_{jk}(Age)$, using discrete step functions.

To this end, consider the partition of the range of possible birth dates into c contiguous intervals, $B_1 = (-\infty, c_1]$, $B_2 = (c_1, c_2]$,...,$B_c = (c_c, \infty)$, where $c_1 < c_2 < ... < c_{c-1}$ are arbitrary breakpoints. Call $\mathcal{B} = \{B_1, B_2..., B_c\}$. Now define the piecewise-constant (on Age) trajectory function, $\lambda_{jk}(Age)$, as

$$\lambda(Age) = \prod_{b \in \mathcal{B}} \left[\lambda_{(b)jk}\right]^{I(Age \in b)} . \tag{2.14}$$

where $0 < \lambda_{(b)jk} < 1$ is a constant for each combination of $b \in \mathcal{B}$, j, k. Note that since $\mathcal{B}$ is a collection of contiguous non-overlapping intervals, this expression is just specifying a step function.

Now we can write the marginal distribution of y_{ijt} as a function of Age using the trajectory function in Equation (2.14)

$$f_{\theta_{j|k}}(y_{ijt}|Age_{it}) \quad = \quad \prod_{b \in \mathcal{B}} \left[{\lambda_{(b)j|k}}^{y} (1 - \lambda_{(b)j|k})^{1-y} \right]^{I(Age_{it} \in b)} .$$

Then, replacing this expression on the generic augmented data likelihood of the model in

Equation (2.6),

$$\begin{aligned} p(y,z|\lambda,g) &\propto \prod_{i=1}^{N}\prod_{j=1}^{J}\prod_{t\in M_i}\prod_{k=1}^{K}\left[g_{ik}f_{\theta_{j|k}}(y_{ijt}|Age_{it})\right]^{I(z_{ijt}=k)} \\ &\propto \prod_{i=1}^{N}\prod_{j=1}^{J}\prod_{t\in M_i}\prod_{b\in\mathcal{B}}\prod_{k=1}^{K}\left[g_{ik}{\lambda_{(b)jk}}^{y_{ijt}}(1-\lambda_{(b)jk})^{1-y_{ijt}}\right]^{I(z_{ijt}=k,Age_{it}\in b)} \end{aligned} \tag{2.15}$$

For the prior distribution for the λ_{jkb} parameters, a natural choice is the beta distribution, that takes values on the interval $(0,1)$. This distribution has the advantage of being conjugate to the binomial distribution and therefore producing extremely simple full conditional distributions (see Erosheva et al. 2007). Additionally, the distribution of Beta$(1,1)$ is the uniform distribution in the interval $(0,1)$, which can be very helpful for expressing prior ignorance about the λ-parameters.

Although in principle we could specify further constraints in the parameters λ (e.g., monotonicity with respect to time, $\lambda_{(b_1)jk} \leq \lambda_{(b_2)jk} \leq \lambda_{(b_3)jk} \dots \leq \lambda_{(b_c)jk}$) I chose not to do it because the whole point of this exercise is to test the effect of having a less constrained trajectory specification.

2.3.1 Estimation

Starting from the augmented data likelihood in Equation (2.15), we can derive the posterior distribution of parameters as

$$p(\alpha, \lambda, g|y, z, x) \propto p(\lambda) \cdot p(\alpha_0) \cdot p(\xi) \left(\prod_{i=1}^{N} p(g_i|\alpha) \right) \\ \times \prod_{i=1}^{N} \prod_{j=1}^{J} \prod_{t \in M_i} \prod_{b \in \mathcal{B}} \prod_{k=1}^{K} \left[g_{ik} {\lambda_{(b)jk}}^{y_{ijt}} (1 - \lambda_{(b)jk})^{1-y_{ijt}} \right]^{I(z_{ijt}=k, Age_{it} \in b)} . \tag{2.16}$$

Analyzing the structure of this expression, we see that the algorithm proposed for the Basic model, with logistic trajectories, in Section 2.2.1 can be applied almost unchanged. We are only left with the task of replacing the steps for sampling the parameters that characterize the extreme trajectories (β parameters in Step 2) with appropriate algorithms for sampling the corresponding parameters, ($\lambda_{(b)jk}$), in the new formulation.

From Equation (2.16) we can easily obtain the full conditional distribution of $\lambda_{(b)jk}$ up to a constant of proportionality

$$p(\lambda_{(b)jk}|...) \propto p(\lambda_{(b)jk}) \times \prod_{i=1}^{N} \prod_{t \in M_i} \left[{\lambda_{(b)jk}}^{y_{ijt}} (1 - \lambda_{(b)jk})^{1-y_{ijt}} \right]^{I(z_{ijt}=k, Age_{it} \in b)} .$$

This expression can be further simplified by defining the set of indexes $\Xi = \Xi_{bjk}^{(it)} = \{(i, t) :$

$Age_{it} \in b, z_{ijt} = k\}$. Thus

$$\begin{aligned}
p(\lambda_{(b)jk}|...) &\propto p(\lambda_{(b)jk}) \times \prod_{\Xi} \lambda_{(b)jk}^{y_{ijt}} (1-\lambda_{(b)jk})^{1-y_{ijt}} \\
&\propto \text{Beta}(\lambda_{(b)jk}|\eta_1, \eta_2) \times \lambda_{jkb}^{\sum_\Xi y_{ijt}} (1-\lambda_{(b)jk})^{\sum_\Xi (1-y_{ijt})} \\
&\propto \lambda_{(b)jk}^{\eta_1 + \sum_\Xi y_{ijt} - 1} (1-\lambda_{(b)jk})^{\eta_2 + \sum_\Xi (1-y_{ijt}) - 1} \\
&\propto \text{Beta}\left[\lambda_{(b)jk} \middle| \eta_1 + \sum_\Xi y_{ijt}, \eta_2 + \sum_\Xi (1-y_{ijt})\right].
\end{aligned}$$

Finally, we can construct an MCMC algorithm for sampling from this model by applying the algorithm for the Basic logistic model in Section 2.2.1, substituting Step 2—sampling of $(\beta_{0jk}, \beta_{1jk})$—by the step:

2'. **Sampling** $\lambda_{(b)jk}$: For each $b \in \mathcal{B}, j \in \{1...J\}$ and $k \in \{1...K\}$ define $\Xi = \{(i,t) : Age_{it} \in b z_{ijt} = k\}$ and sample

$$\lambda_{(b)jk}|... \sim \text{Beta}\left[\eta_1 + \sum_\Xi y_{ijt}, \eta_2 + \sum_\Xi (1-y_{ijt})\right].$$

2.4 Comments

I have described a method to handle multivariate trajectories of the probability of a binary outcome as a function of time, while simultaneously soft-classifying the population with respect to a small number classes. The model can be understood as a blend of the Grade of Membership (Woodbury et al., 1978; Erosheva et al., 2007) and the Multivariate Latent Trajectory model (Connor, 2006). I have presented the model in general terms and then specialized it to the task of modeling disability patterns in the NLTCS, proposing a full

Hierarchical Bayes specification. I have also developed estimation techniques based on MCMC posterior sampling to approximate the posterior distribution of the parameters.

Besides priors, these models require the selection of specific extreme trajectory functions to complete the specification, with a particular application in mind. Specifically for the NLTCS case, I have selected the s-shaped logistic curve, proposed in Connor (2006) as a reasonable parametric alternative; and another, less constrained one, based on step functions, for testing purposes. Other specifications are certainly possible, including less restricted ones. An interesting alternative is the non-parametric splines framework proposed in Dimatteo et al. (2001).

In the specialization for the analysis of the NLTCS, the estimation algorithm relies on a strong Missing At Random (MAR) assumption, in order to handle the missing values (Rubin, 1976). This is particularly important since, even though the NLTCS individual disability outcomes can be understood as a rectangular $J \times T$ array, by the survey's own design a large number of individuals have one or more missing measurements. Although the MAR assumption is reasonable for some of the sources of missingness, it is a highly suspicious assumption when it comes to missingness due to death, as death is an event that is likely to be correlated with the disability responses. We return to this and related issues in Chapter 6.

A possible objection to the use of continuous latent trajectories to model disability progression in the NLTCS is that the long inter-measurement intervals (5 years), coupled with the fact that the observations are discrete, could provide too little information for accurate estimation. While in fact, considered in isolation, a few discrete measurements (between 1 and 6 in the NLTCS), spaced every five years provide very limited information, we have to bear in mind that the trajectories are estimated conditionally on *personal time*, as opposed to chronological time. This means that at each measurement, determined by a point in the chronological time continuum, we have several subjects that are at distinct points of their personal time (i.e. have different ages). Therefore, the *combined* pool of data provides us with plenty of points to perform very detailed reconstruction of trajectories.

Figure 2.1 illustrates this situation. The left panel in Figure 2.1 presents a (hypothetical) sample of four individuals that have the exact same underlying trajectory but were born in different years. They entered the sample at different points in time (waves) and have a different number of measurements, taken at each wave where they participated. This situation is represented by four trajectory curves that differ only by their location. Vertical lines represent the points in time where the survey waves took place; asterisks, '(*)', the waves at which each of these individuals first entered the sample and; the points over the curves, the measurements registered by the survey. Note that, to help with the visualization, in the plot the "measurements" are taken from the actual underlying probabilities, instead of the realizations. Also note that not every individual has measurements in each of the waves. The center panel shows the effect of translating the chronological time into personal time (age): a "realignment effect", where each measurement now lies on a comparable time line. Finally (right panel), combining all the measurements, we see that we actually have several data points to reconstruct the underlying trend.

This effect can be seen as an advantage of this particular approach to modeling the NLTCS, that enables the estimation of smooth and interpretable trajectories from very sparse individual information, but also exposes one of its main weaknesses: considering individuals from different generations as essentially exchangeable. Whether this is a reasonable assumption or not is a debatable issue. What is certain, though, is that this approach limits

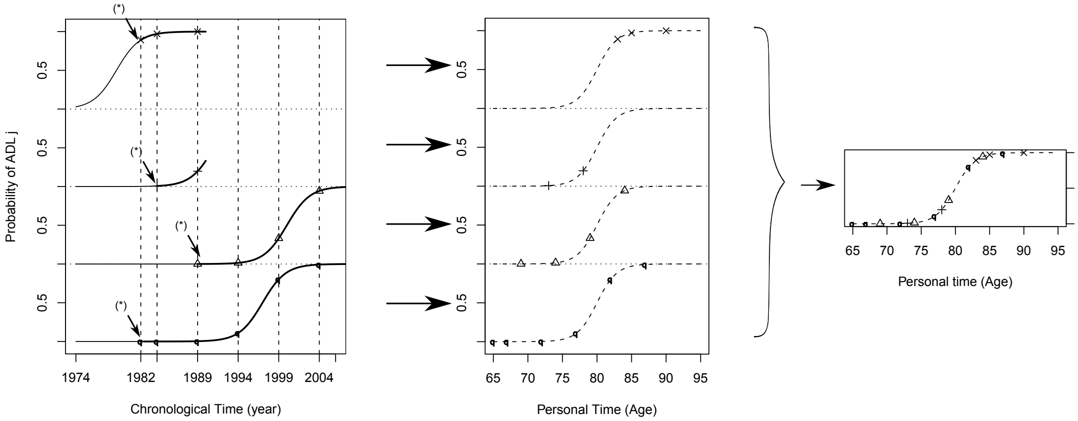


Figure 2.1: "Realignment effect" when translating the discrete measurements taken from different individuals to a common age scale.

apriori our ability to attack one of the main issues in the NLTCS analysis, namely the investigation of the changes from one generation to another (e.g. "Are individuals from younger generations acquiring disabilities differently than older ones?"). We will return to this question in Chapter 6.

2.A Appendices

2.A.1 Computational Issues

Sampling from Gamma distributions with small shape parameter

All the algorithms presented in this Section (and all of the ones in the following sections) have in common Step 3, where we are required to sample, N values g_i from Dirichlet distributions. The usual way of obtaining variates g from a Dirichlet$(\alpha_1, \alpha_2, ..., \alpha_K)$ distribution is through the following algorithm (Devroye, 1986):

1. For $k = 1...K$, sample $\alpha_k \overset{indep}{\sim} \text{Gamma}(\alpha_k, 1)$.

2. Compute $\alpha_0 = \sum_{k=1}^{K} \alpha_k$.

3. Return $g = (\frac{\alpha_1}{\alpha_0}, \frac{\alpha_2}{\alpha_0}, ..., \frac{\alpha_K}{\alpha_0})$.

During the normal execution of the MCMC samplers it is frequent that at some regions we have to sample from Gamma distributions with very small shape parameter (e.g. $\alpha_k < 0.01$). Since these distributions have most of their mass concentrated towards the origin, most samplers will return several 0 values, due to numerical underflow. This can lead to serious computational problems in algorithms where we depend on the positivity of these samples. For instance, during the execution of the sampler for the basic model, Step 4 requires the computation of the product, $\prod_{i=1}^{N} g_{ik}$, which will be driven to 0 if any of these values has been sampled as 0.

One way to overcome this problem is to make use of the following fact (Robert and Casella, 2004):

$$\mathcal{L}[\text{Gamma}(\alpha, 1)] = \mathcal{L}\left[U[0, 1]^{1/\alpha} \cdot \text{Gamma}(1 + \alpha, 1)\right].$$

Then we sample directly the log transformed gamma variates, $\log(Y) \sim \log[\text{Gamma}(\alpha, 1)]$, through the following two-step algorithm:

1. Sample $U \sim U[0, 1]$.

2. Sample $G \sim \text{Gamma}(1 + \alpha, 1)$.

3. Return
$$lY = \frac{\log U}{\alpha} + \log G.$$

Values sampled with this algorithm will be such that $\exp lY \sim \text{Gamma}(\alpha, 1)$. These variates can be used directly to generate log transformed Dirichlet variates and it is recommended to store them in this form for future use, in order to avoid their degradation by floating point numerical rounding.

This method has the advantage that, instead of having to sample from a $\text{Gamma}(\alpha, 1)$ distribution where the shape parameter $\alpha \ll 1$ would likely produce a numerical underflow to 0, we just sample from the much more numerically stable $\text{Gamma}(\alpha + 1, 1)$ and $\text{U}(0, 1)$ distributions, and preserve the sampled variates in log transformed form.

Efficient sampling of β parameters

After the production of the $N \times J \times T$ samples of z_{ijt}, the most costly part of the algorithm for sampling from the basic logistic model is the computation of the rejection ratio for the Metropolis step in the sampling of parameters β_{jk}. In particular, the expression

$$\begin{aligned} r_{jk} &= \prod_{\Xi_{jk}} \left[\frac{1 + \exp\left(\beta_{0jk} + \beta_{1jk} Age_{it}\right)}{1 + \exp\left(\beta^*_{0jk} + \beta^*_{1jk} Age_{it}\right)} \right] \\ &\quad \times \exp\left[-\frac{\beta^{*2}_{0jk} - \beta^2_{0jk}}{2\sigma_0^2} + \left(\beta^*_{0jk} - \beta_{0jk}\right) \left(\frac{\mu_0}{\sigma_0} + \sum_{\Xi_{jk}} y_{ijt} \right) \right] \\ &\quad \times \exp\left[-\frac{\beta^{*2}_{1jk} - \beta^2_{1jk}}{2\sigma_1^2} + \left(\beta^*_{1jk} - \beta_{1jk}\right) \left(\frac{\mu_1}{\sigma_1} + \sum_{\Xi_{jk}} y_{ijt} Age_{it} \right) \right] \end{aligned}$$

consists of a main multiplication and also contains two sub-expressions that have to be computed iterating over the indexes (i, t). Each complete cycle represents $N \times \overline{(N_i)} = \sum_i N_i$ iterations. A naive implementation of this computation, iterating over each β_{jk} and computing its corresponding r_{jk} would require $J \times K \times 3 \times \sum_i N_i$ iterations.

A simple observation of the structure of these operations however tells us that, if we consider all r_{jk} indexes simultaneously, at any given combination of indexes (i, j, t) there is only one value of r_{jk} being actually computed, since there is only one possible value of k that will make $k = z_{ijt}$. In other words, $\Xi_{jk} \cap \Xi_{jk'} = \emptyset$ for $k \neq k'$. This means that we can exploit the special structure of this problem to reduce the number of iterations by computing all the r_{jk} factors simultaneously using the following algorithm:

```
for all $j = 1, ..., J$ and $k = 1...K$ do
    $a'_{jk} \leftarrow 1$, $a''_{jk} \leftarrow 0$, $a'''_{jk} \leftarrow 0$
end for
for all $i = 1...N$, $j = 1...J$, $t \in M_i$ do
    $k \leftarrow z_{ijt}$
    $a'_{jk} \leftarrow a'_{jk} \times \frac{1+\exp\left[\left(\beta_{0jk}+\beta_{1jk} Age_{it}\right)\right]}{1+\exp\left[\left(\beta^*_{0jk}+\beta^*_{1jk} Age_{it}\right)\right]}$
    $a''_{jk} \leftarrow a''_{jk} + y_{ijt}$
    $a'''_{jk} \leftarrow a'''_{jk} + Age_{it} \cdot y_{ijt}$
end for
for all $j = 1...J$ and $k = 1...K$ do
    $r_{jk} \leftarrow a'_{jk}$
    $r_{jk} \leftarrow r_{jk} \times \exp\left[\frac{1}{2\sigma_0^2}(\beta^{*2}_{0jk} - \beta^2_{0jk}) - \frac{1}{2\sigma_1^2}(\beta^{*2}_{1jk} - \beta^2_{1jk})\right]$
    $r_{jk} \leftarrow r_{jk} \times \exp\left[(\beta^*_{0jk} - \beta_{0jk})\left(\frac{\mu_0}{\sigma_0} + a''_{jk}\right)\right]$
    $r_{jk} \leftarrow r_{jk} \times \exp\left[(\beta^*_{1jk} - \beta_{1jk})\left(\frac{\mu_1}{\sigma_1} + a'''_{jk}\right)\right]$
end for
return $r_{jk}$ for all $j = 1...J$ and $k = 1...K$
```

Given that at each sampling step no β_{jk} depends on the other $\beta_{j'k'}$s, this algorithm will in fact compute all the $J \times K$ needed rejection ratios, r_{jk}, but in just $J \times (\sum_i N_i + K)$ iterations, decreasing the number of needed iterations by a factor of approximately $3 \times K$ respect to the naive implementation. This same observation applies almost directly to the basic discrete model from Section 2.3.

Chapter 3

Applying the Basic Model to the NLTCS data

In order to illustrate the methods proposed in Chapter 2 I apply them to the NLTCS data described in Section 1.1. I fit models with $K = 2, 3, 4$ and 5 extreme profiles and estimated the posterior distribution of the parameters. Finally, I comment on how to interpret the results and on the practicalities of the application of the algorithms.

3.1 Preparation and Preliminary Issues

3.1.1 Data

The dataset used for these computations includes all measurements in all waves ($T = 6$) of the subjects that received the screener in at least one of the 1982, 1984, 1989, 1994 or

1999 waves. New subjects that entered during the 2004 wave had to be excluded because of lack of information about their dates of birth and death. Similarly, all subjects that in 1982 were institutionalized were excluded because of lack of ADL recording. The final number of subjects was then $N = 39,323$. The response vector includes the six ($J = 6$) binary coded ADLs: Eating ($j = 1$), Dressing ($j = 2$), Toileting ($j = 3$), Getting In or Out of Bed ($j = 4$), Inside Mobility ($j = 5$) and Bathing ($j = 6$). The age has been computed substracting the date of birth from the exact date of the interview. I have converted the ages from days to years considering the equivalence 1 year = 365 days. To facilitate the computations and prior specification, I have also re-centered the age values by substracting 80 years.

3.1.2 Prior distributions

I have set the prior distribution for $\alpha = \alpha_0 \cdot \xi$ as independent $\alpha_0 \sim \text{Gamma}(1, 5)$ and $\xi \sim \text{Dirichlet}(\mathbf{1}_K)$. This prior expresses the notion of complete ignorance about the relative importance of the extreme profiles in the population, and preference for smaller values of the concentration parameter, α_0. The reasons behind the last choice are mostly interpretative: As I explained in Chapter 2, Section 2.2, a Dirichlet distribution with small values of α_0 will produce individual realizations that are closer to one particular vertex of the simplex, with influence of the other vertices and, as α_0 goes all the way down to 0, a degenerate discrete distribution over the vertices. This arrangement will allow us to talk about a "dominant profiles" that are influenced by the others, easing the interpretation of the results while still allowing the mixed membership apparatus to handle the extra heterogeneity. In this way, I do not consider this choice of prior distribution to be an expression of prior knowledge but as a *modeling decision*, in the same sense as the selection of the parametric trajectory curves is.

For choosing the prior distribution of β_0 and β_1 we first may notice that the greatest slope of any extreme trajectory will be achieved at the inflection point of the logistic curve, at $Age = 0$, with value $\beta_1/4$ (recall that ages re centered at 80 years). The prior distribution of β_1 then depends on the rate of change of the underlying probability of disability. I choose $\beta_{1jk} \sim N(\mu = 0, \sigma^2 = 100)$. This distribution could be considered very diffuse as, considering that about 95% of its values will be within 2 standard deviations from 0, most of the slopes will range from a flat trajectory ($\beta_{1jk} = 0$) to changes of up to 100% in just 20 days ($|\beta_{1jk}| = 2\sigma$).

Similarly, I choose the prior distribution of β_0 to be $\beta_{0jk} \sim N(0, 100)$ based on the following observations. If we consider that the center of the curve is at its inflection point, the trajectory curve is centered around the point $Age = -\beta_0/\beta_1$. Then, since the proposed prior specification for β_0 is centered around 0, small deviations in the location of the curve from $Age = 0$ are already accounted for. Considering that typical individual ages in the NLTCS range between 65 and 100 years ($Age \in [-15, 20]$), a reasonable large deviation in location from $Age = 0$, $|\beta_0/\beta_1|$, to consider is 20 years. Thus, for an equally reasonably large value of $|\beta_1| = 0.4$ (roughly a change in probability of disability from 0 to 100% in just one month), the greatest magnitude of β_0 that we need to consider to achieve such an offset of 20 years is $|\beta_0| = 8$. Since the above prior specification has most of its mass in the interval $[-20, 20]$ I consider it be diffuse over the domain of interest.

Extreme Profiles	σ^*_α
$K = 2$	0.026
$K = 3$	0.011
$K = 4$	0.009
$K = 5$	0.007

Table 3.1: Values of proposal standard deviation σ^*_α for basic model with continuous logistic trajectories.

3.1.3 Execution of the Algorithm

For the first tests of my algorithm, I took random samples from the prior distribution as initial values. After several trials with the chains converging to essentially the same posterior distributions, I set the starting values to the following: $\alpha_0 = 1/3$, $\beta_{0jk} = 0$, $\beta_{1jk} = 0$ and $\xi_k = 1/K$.

I chose the tuning parameters to keep the acceptance rates at around 20%. This was especially difficult with the vector α, that turned out to be very sensitive to small variations of the tuning parameter. The final values for the tuning parameter σ^*_α are shown in Table 3.1. These specific values achieved acceptance rates of $20\%, 19\%, 15\%$ and 20% for $K = 2, 3, 4$ and 5, respectively. For the β parameters, I used $\sigma^*_0 = 0.2$ and $\sigma^*_1 = 0.02$ for every pair $(\beta_{0jk}, \beta_{1jk})$, achieving acceptance rates between 5% and 30%, for distinct combinations of the indexes (j, k), with most of the values around 15%. While it is certainly possible to use different proposals for each pair jk, the use of a single distribution was a compromise solution that has the advantage of dispensing us from the work of manually tune $2 \times J \times K$ different proposal distributions.

These chains converge quite quickly, after around 15,000 iterations, but exhibit a rather high autocorrelation. For this reason, from 100,000 long runs, I discarded the first $20,000$

iterations as a burn-in period and subsampled it, taking a sample every 5 samples and discarding the rest (Robert and Casella, 2004). Although an even less frequent subsampling was possible and would have reduced the autocorrelation of the chain, the differences in the final estimates turn out to be negligible.

3.2 Results

In Tables 3.2 to 3.5 I have arranged the main estimated parameters (marginal posterior means) of the models with $K = 2, 3, 4$ and 5 extreme profiles. Along with the parameters of the model I have also computed the quantity

$$Age_{1/2,jk} = -\beta_{0jk}/\beta_{1jk} + 80,$$

defined as the age where the trajectory for ADL-j in extreme profile k crosses the 0.5 probability threshold. This quantity is related to the "age of onset" of disability and is useful to provide insight into the temporal characteristics of the extreme profiles. Additionally, since this model (and any latent class model) is invariant to permutations of the class labels, I have reordered the extreme profiles according to the estimates of parameter ξ_k.

There are several things to note in these results:

1. Estimates for the parameter α_0 are consistently small for all models. This was expected since the prior distribution of α_0, Gamma$(1, 5)$, was in fact already expressing strong apriori preference for small values of α_0. We see, however, that 1) these estimates are strongly data-driven, meaning that the posterior distribution is

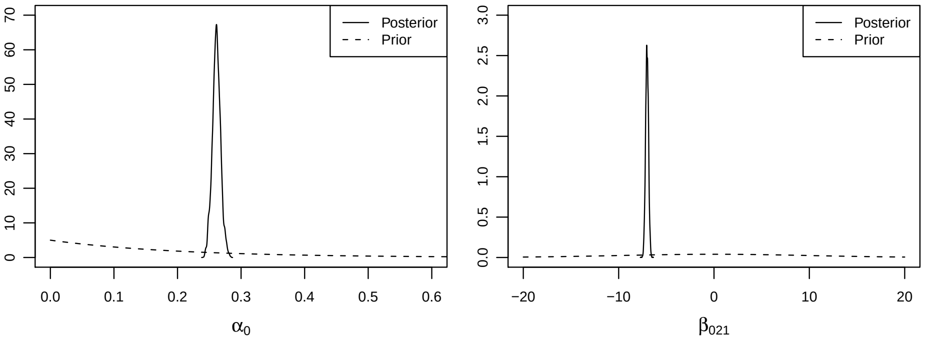
much more concentrated around a particular value than the prior and 2) that the estimated value of α_0 still yields a distribution over the simplex, spread enough to allow for non-trivial individual-level mixing of extreme trajectories. I comment more on these two points later in this section.

2. The posterior standard deviation of all estimates is very small. This was expected because of the abundance of data. In fact, we can obtain essentially the same estimates (with higher dispersion) using just a fraction (e.g. $N = 6000$) of the available data.

3. For each extreme profile, $k = 1, 2..., K$, the sequence of ADLs obtained after sorting them according to the estimate of parameter $Age_{1/2,jk}$ is exactly the same: inside mobility ($j = 5$), toileting ($j = 3$), dressing ($j = 2$), bathing ($j = 6$), getting in and out of bed ($j = 4$) and eating ($j = 1$). This is quite meaningful, as it informs us of a seemingly universal temporal ordering among the different manifestations of disability that resembles an intuitive ordering of tasks in terms of difficulty.

4. All the parameters β_1—except for profile $k = 5$ in the model with $K = 5$ extreme profiles—are positive. This will produce, at least for models with less than five extreme profiles, sound increasing trajectories, that exhibit a progressive increase in the probability of acquiring a disability as the subject ages.

Figure 3.1 shows kernel density estimates of the posterior distributions of α_0 and β_{021}, together with their corresponding prior distribution, to illustrate the relationship between priors and posteriors. Compared with their prior distributions, the posteriors are concentrated around particular values. This is the case for all the parameters and hyperparameters of the model. This fact is particularly interesting in the case of α_0, whose



(a) Prior versus posterior α_0 (b) Prior versus posterior β_{021}

Figure 3.1: Prior and posterior distributions for parameters α_0 and β_{021} for basic model with $K = 3$. The prior distribution for α_0 is Gamma$(1, 5)$ and for β_{0jk} is $N(0, 100)$.

prior distribution was already very constraining. What this comparison illustrates is that, although the prior distribution expresses preference for values of α_0 to be near the origin, the data dominates in the posterior estimation.

Figures 3.2 to 3.5 show plots of the extreme trajectories, for every extreme profile (rows) and ADL (columns), for models with $K = 2, 3, 4$ and 5 extreme profiles. These plots, first introduced by Connor (2006), provide an intuitive idea of the underlying typical trends in the population and are a good complement to the tabulation of posterior summaries. Again we can observe some salient features:

1. The first extreme profile in all models (after sorting the extreme profiles decreasingly according to the value of ξ_k) exhibit aging progressions where people remain basically healthy until approximately the age of 90, and then start to experience an increase in their chances of acquiring a disability.

2. As we consider models with different number of extreme profiles, with extreme profiles sorted according to the posterior estimate of the component ξ_k, from $K = 2$

(Figure 3.2) ,to $K = 5$ (Figure 3.5), we observe that the extreme trajectories form what essentially seems to be a decreasing gradation on the age of onset of disability between the first extreme profile (essentially healthy) and the last one (very early onset of disability). This is quite interesting since the sequence of extreme profiles has been organized only based on their relative importance in the population (parameter ξ_k). This points to the existence of an inverse relationship between the relative preeminence of each extreme profile in the population and the age of onset of disability that it is expressing. Exceptions to this tendency are the trajectory for ADL 1 (eating) in extreme profile $k = 4$ in the model with $K = 4$ (Figure 3.4) and the whole extreme profile $k = 5$ in the model with $K = 5$ extreme profiles (Figure 3.5).

3. Trajectories in profile $k = 5$ in model with $K = 5$ (Figure 3.5) show a somewhat erratic behavior. In particular many of them exhibit decreasing chances of disability as individuals age. This is certainly counterintuitive and goes against the intended interpretation of the model. Closer attention to the characteristics of the profile, in Table 3.5, reveals that its relative preeminence in the population is very small, with $\xi_5 = 0.058$ (c.f. $\xi_1 = 0.496$ for profile $k = 1$). This suggests that there might actually be very few individuals in the sample who are close to this profile. We will explore this issue further in the Section 3.3 and in Chapter 6.

Since decreasing trajectories go against the intended interpretation of the model, a natural way of dealing with this situation is simply to constrain the specification so that those cases are excluded by design. This can be achieved by modifying the prior distributions of the slope parameters, β_1, replacing them with distributions with support in positive real values. I have implemented such a variation (see Appendix 3.A for details on the implementation), but the results pointed in the same direction: for models with $K \geq 5$ extreme profiles, the slopes of several trajectories of the less predominant profiles are zero or close to zero. The rest of the estimates were very similar to the ones obtained with the less constrained specification. Since nothing seems to be gained from this specification, I do not consider it further.

We can use estimates of the individual-level posterior membership vectors to construct instances of the estimated individual-level trajectory curves, by weighting each curve by its corresponding estimated membership vector component, g_{ik}, as specified in Equation (2.1) in Section 2, for each individual. Figures 3.6, 3.7 and 3.8 show 100 of these individual level trajectory curves, overlaid with the extreme trajectories as a reference, for all the fitted models. From these pictures we can form a good idea of how this method is handling the individual-level heterogeneity. We see that, although most of the curves cluster on the vicinity of one of the extreme trajectories, as expected from the low posterior values of α_0, several of them lie somewhere in the middle, with trajectories that are the product of the interaction of the extreme profiles. In this way, we can see that these models have enough flexibility to accommodate several trajectories that do not lie exactly in one of the extreme profiles but somewhere in between.

This last point deserves further attention. Looking at the individual level trajectories for $K = 3$ (Figure 3.7) and $K = 4$ (Figure 3.8), besides the clustering of individual trajectories close to the extreme trajectories (expected from the estimated values of α_0), we see that other “clusters” appear to have emerged in the posterior estimates. This means that the models are in fact picking up unexpected features of the data while keeping the number of extreme profiles to a minimum. If we were estimating with a model based on full-membership or latent class approaches, like the ones in Connor (2006), without the help of the mixed-membership device, we would have needed at least an extra class to make room to those individuals that compose the emerging cluster.

Table 3.6 presents observed and posterior predictive counts from the 2^6 possible response patterns of ADLs, aggregated by age range. As a way of comparison it also presents the X^2 statistic computed from the resulting $2^6 \times 5$ contingency table for all the fitted models. We can observe that, in terms of prediction, models perform quite well, with the possible exception of $K = 2$. The only problematic cells seem to be those corresponding to the people who report disabilities in all ADLs simultaneously ('111111' in Table 3.6), that tend to be consistently underestimated in all models. We can also see that the cells corresponding to no disabilities ('000000' in Table 3.6) tend to have consistently good estimates. This is quite interesting since, as reported by Erosheva et al. (2007), regular GoM models tend to perform poorly when it comes to fitting non disabled states.

3.3 Using the Discrete Trajectories Variant

As I mentioned in Sections 2.2 and 2.3, in Chapter 2, the choice of the particular trajectory function used in the estimations in the previous section is arbitrary. While it has produced estimates of the extreme trajectories that are both reasonable and sound (up to $K = 4$), one might wonder what is the effect of the strong parametric assumptions implied, in particular monotonicity and smoothness, and what would happen if those assumptions are relaxed.

In order to assess this effect I fit the model to the same NLTCS data using the Discrete Trajectories variant of the basic model, introduced in Section 2.3 of Chapter 2. Remember

Parameter	Estimate
α_0	0.328 (0.007)

Parameter	ADL(j)	Estimate Extreme Profile-k (sd)			
		$k=1$		$k=2$	
ξ	–	0.824	(0.002)	0.176	(0.002)
β_{0**}	1 (EAT)	-6.496	(0.121)	-0.494	(0.026)
	2 (BED)	-4.861	(0.063)	1.736	(0.053)
	3 (MOB)	-3.789	(0.042)	2.843	(0.084)
	4 (DRS)	-5.518	(0.082)	0.945	(0.039)
	5 (BTH)	-3.163	(0.032)	3.298	(0.097)
	6 (TLT)	-4.710	(0.059)	1.292	(0.045)
β_{1**}	1 (EAT)	0.293	(0.008)	0.136	(0.003)
	2 (BED)	0.312	(0.005)	0.218	(0.005)
	3 (MOB)	0.289	(0.004)	0.273	(0.007)
	4 (DRS)	0.310	(0.006)	0.182	(0.004)
	5 (BTH)	0.247	(0.003)	0.295	(0.008)
	6 (TLT)	0.280	(0.005)	0.204	(0.005)
$Age_{1/2}$	1 (EAT)	102.171	(0.294)	83.641	(0.236)
	2 (BED)	95.577	(0.130)	72.038	(0.144)
	3 (MOB)	93.128	(0.107)	69.589	(0.132)
	4 (DRS)	97.818	(0.173)	74.824	(0.155)
	5 (BTH)	92.823	(0.120)	68.810	(0.124)
	6 (TLT)	96.831	(0.160)	73.663	(0.144)

Table 3.2: Posterior means for parameters of interest for basic model with $K = 2$ extreme profiles. Numbers between parenthesis are posterior standard deviations.

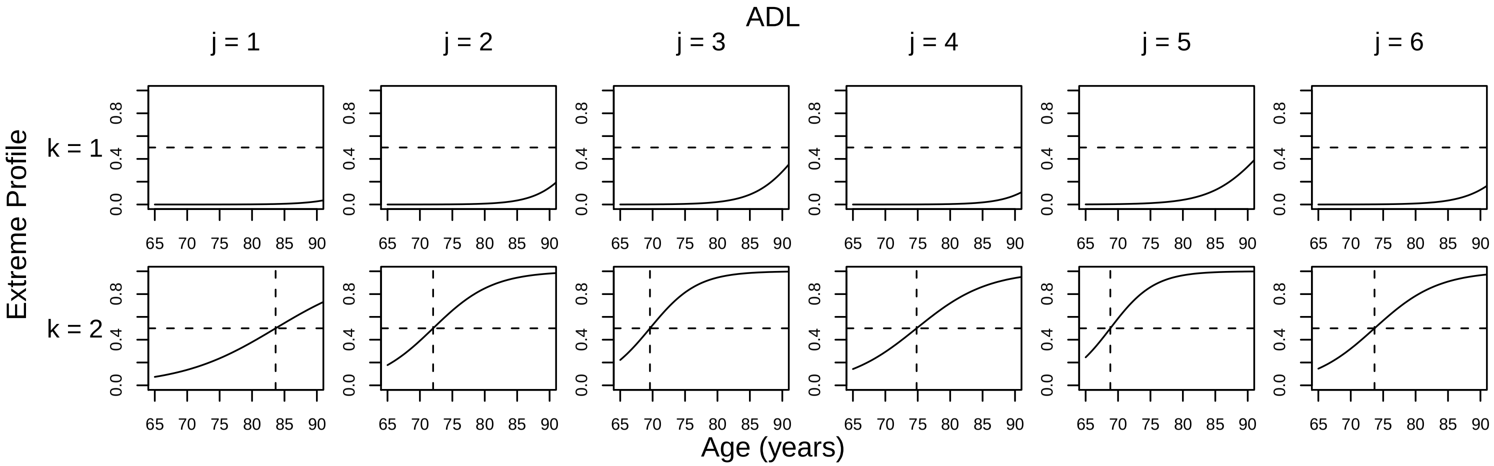


Figure 3.2: Extreme trajectories of disability over time for each ADL and extreme profile ($K = 2$) Dashed lines indicate the point where the probability of acquiring a disability in ADL j reaches 50%.

Parameter	Estimate
α_0	0.261 (0.006)

Parameter	ADL(j)	Estimate Extreme Profile-k (sd)					
		$k=1$		$k=2$		$k=3$	
ξ	–	0.645	(0.004)	0.251	(0.004)	0.104	(0.002)
β_{0**}	1 (EAT)	-8.843	(0.322)	-3.100	(0.060)	-0.074	(0.046)
	2 (BED)	-7.022	(0.151)	-1.737	(0.055)	3.559	(0.151)
	3 (MOB)	-5.332	(0.096)	-0.757	(0.046)	5.734	(0.274)
	4 (DRS)	-7.885	(0.218)	-2.249	(0.055)	2.028	(0.084)
	5 (BTH)	-4.452	(0.071)	-0.229	(0.039)	6.210	(0.304)
	6 (TLT)	-6.587	(0.145)	-1.768	(0.049)	2.491	(0.100)
β_{1**}	1 (EAT)	0.357	(0.017)	0.347	(0.009)	0.104	(0.006)
	2 (BED)	0.394	(0.010)	0.551	(0.013)	0.289	(0.013)
	3 (MOB)	0.348	(0.007)	0.520	(0.012)	0.422	(0.022)
	4 (DRS)	0.390	(0.013)	0.462	(0.011)	0.202	(0.008)
	5 (BTH)	0.295	(0.006)	0.425	(0.009)	0.442	(0.024)
	6 (TLT)	0.337	(0.009)	0.474	(0.011)	0.233	(0.009)
$Age_{1/2}$	1 (EAT)	104.768	(0.462)	88.933	(0.172)	80.725	(0.477)
	2 (BED)	97.824	(0.174)	83.156	(0.091)	67.674	(0.179)
	3 (MOB)	95.332	(0.139)	81.457	(0.084)	66.399	(0.155)
	4 (DRS)	100.210	(0.235)	84.873	(0.105)	69.959	(0.197)
	5 (BTH)	95.107	(0.155)	80.539	(0.093)	65.940	(0.164)
	6 (TLT)	99.538	(0.231)	83.731	(0.097)	69.315	(0.181)

Table 3.3: Posterior means for parameters of interest for basic model with $K = 3$ extreme profiles. Numbers between parenthesis are posterior standard deviations.

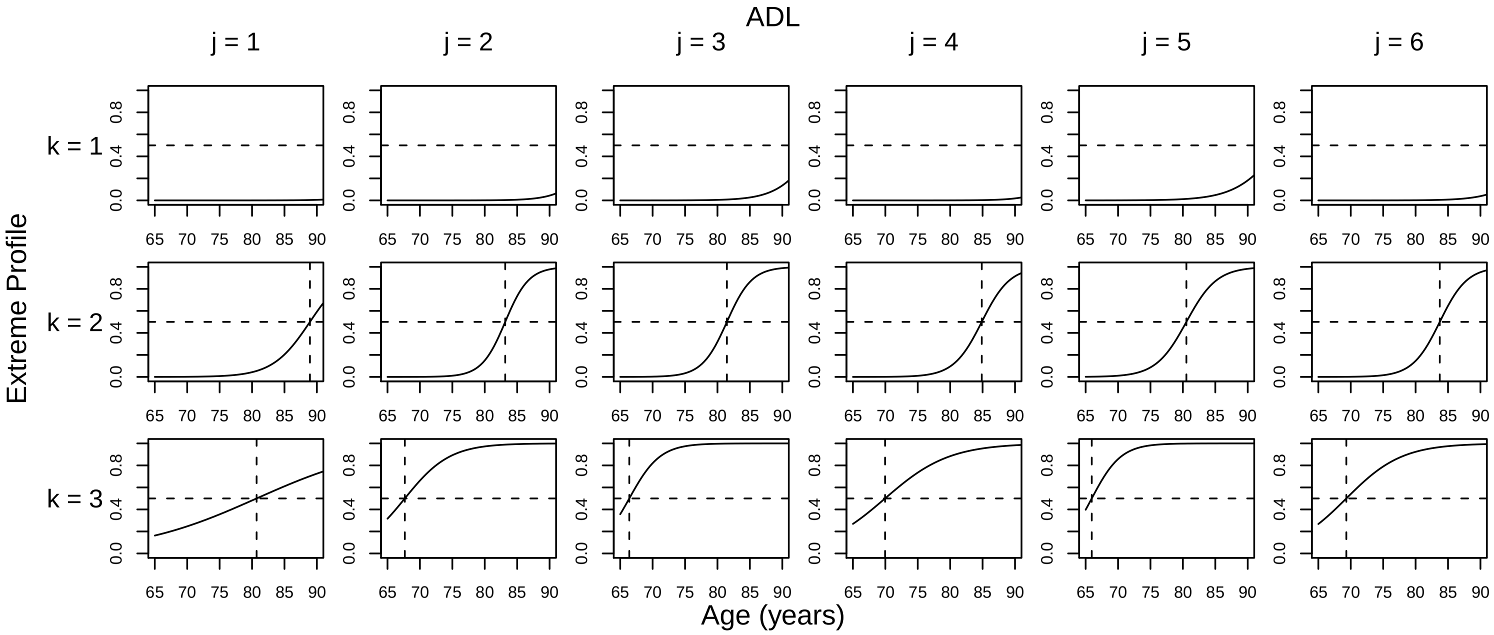


Figure 3.3: Extreme trajectories of disability over time for each ADL and extreme profile ($K = 3$) Dashed lines indicate the point where the probability of acquiring a disability in ADL j reaches 50%.

Parameter	Estimate
α_0	0.237 (0.006)

Parameter	ADL(j)	Estimate Extreme Profile-k (sd)							
		$k = 1$		$k = 2$		$k = 3$		$k = 4$	
ξ	–	0.540	(0.005)	0.259	(0.004)	0.124	(0.003)	0.078	(0.002)
β_{0**}	1 (EAT)	-10.506	(0.633)	-5.109	(0.130)	-1.281	(0.064)	-0.332	(0.058)
	2 (BED)	-7.975	(0.219)	-4.745	(0.153)	0.897	(0.082)	4.727	(0.327)
	3 (MOB)	-6.035	(0.138)	-3.281	(0.111)	1.627	(0.081)	7.986	(0.634)
	4 (DRS)	-9.310	(0.378)	-5.164	(0.152)	0.158	(0.080)	2.003	(0.152)
	5 (BTH)	-5.028	(0.116)	-2.370	(0.078)	1.898	(0.078)	8.998	(0.827)
	6 (TLT)	-7.514	(0.225)	-4.353	(0.124)	0.594	(0.077)	2.520	(0.179)
β_{1**}	1 (EAT)	0.407	(0.030)	0.449	(0.015)	0.416	(0.015)	0.038	(0.007)
	2 (BED)	0.418	(0.013)	0.754	(0.025)	0.634	(0.022)	0.320	(0.026)
	3 (MOB)	0.364	(0.010)	0.690	(0.021)	0.547	(0.018)	0.525	(0.048)
	4 (DRS)	0.431	(0.020)	0.663	(0.021)	0.560	(0.020)	0.145	(0.015)
	5 (BTH)	0.305	(0.008)	0.576	(0.017)	0.461	(0.012)	0.580	(0.062)
	6 (TLT)	0.354	(0.013)	0.639	(0.020)	0.549	(0.018)	0.182	(0.017)
$Age_{1/2}$	1 (EAT)	105.818	(0.589)	91.394	(0.196)	83.086	(0.195)	89.155	(3.138)
	2 (BED)	99.097	(0.213)	86.297	(0.098)	78.586	(0.114)	65.200	(0.329)
	3 (MOB)	96.570	(0.175)	84.754	(0.095)	77.022	(0.119)	64.770	(0.280)
	4 (DRS)	101.626	(0.305)	87.794	(0.113)	79.720	(0.139)	66.178	(0.563)
	5 (BTH)	96.486	(0.191)	84.119	(0.095)	75.886	(0.133)	64.454	(0.323)
	6 (TLT)	101.210	(0.307)	86.815	(0.107)	78.918	(0.128)	66.106	(0.481)

Table 3.4: Posterior means for parameters of interest for basic model with $K = 4$ extreme profiles. Numbers between parenthesis are posterior standard deviations.

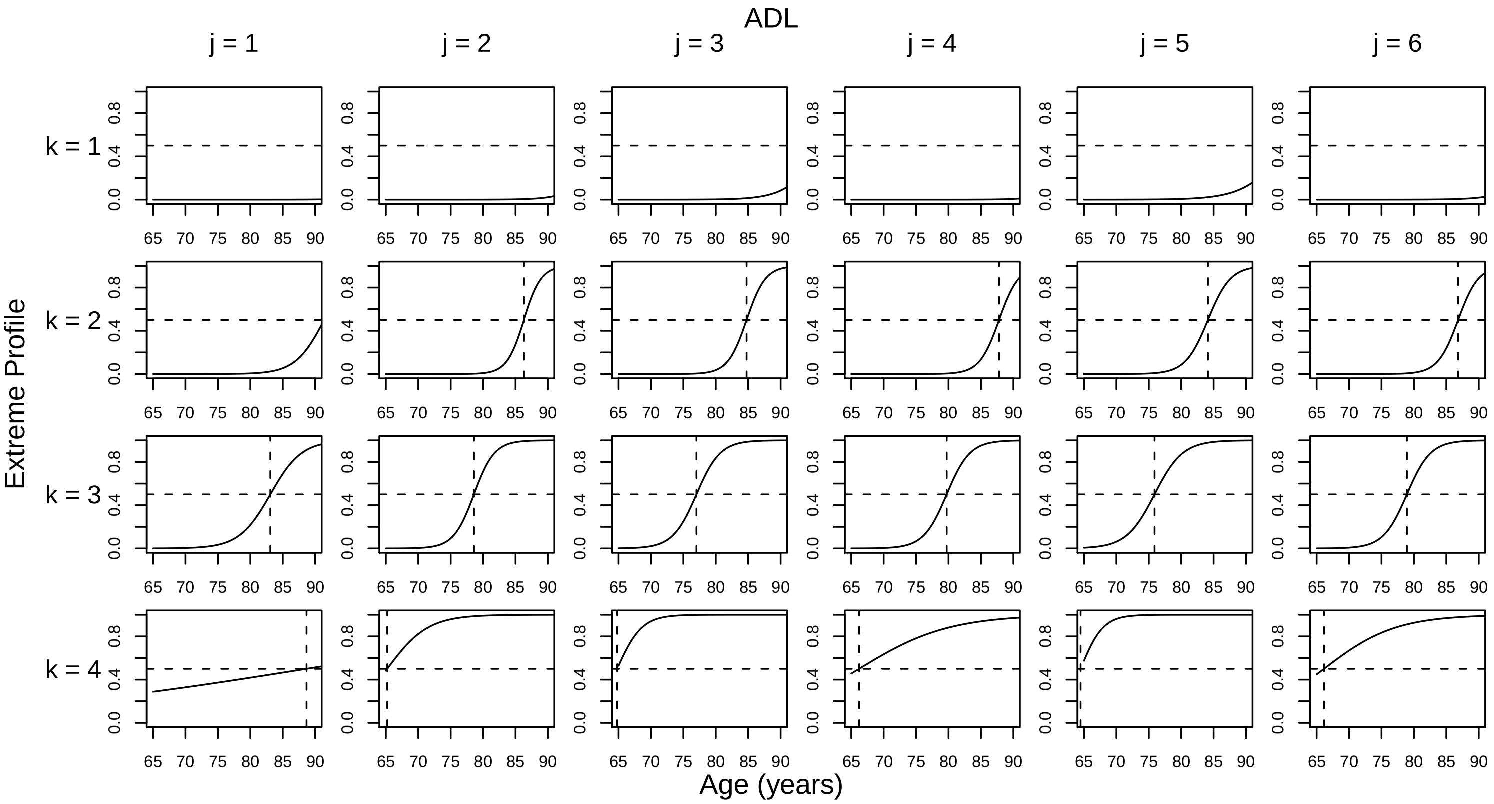


Figure 3.4: Extreme trajectories of disability over time for each ADL and extreme profile ($K = 4$) Dashed lines indicate the point where the probability of acquiring a disability in ADL j reaches 50%.

Parameter	Estimate
α_0	0.235 (0.005)

Parameter	ADL(j)	Estimate Extreme Profile-k (sd)									
		$k=1$		$k=2$		$k=3$		$k=4$		$k=5$	
ξ	–	0.496	(0.007)	0.244	(0.006)	0.128	(0.003)	0.074	(0.002)	0.058	(0.001)
β_{0**}	1 (EAT)	-11.114	(0.650)	-5.868	(0.174)	-2.538	(0.082)	0.503	(0.099)	-1.578	(0.080)
	2 (BED)	-9.904	(0.364)	-6.969	(0.268)	-1.131	(0.104)	4.509	(0.230)	1.571	(0.125)
	3 (MOB)	-7.537	(0.222)	-4.791	(0.182)	0.009	(0.082)	5.114	(0.242)	31.422	(5.124)
	4 (DRS)	-10.971	(0.515)	-7.139	(0.262)	-1.650	(0.093)	3.632	(0.201)	-0.090	(0.075)
	5 (BTH)	-6.156	(0.189)	-3.371	(0.118)	0.593	(0.069)	5.080	(0.210)	31.188	(5.260)
	6 (TLT)	-8.559	(0.305)	-5.930	(0.203)	-1.123	(0.088)	4.120	(0.210)	0.376	(0.069)
β_{1**}	1 (EAT)	0.473	(0.032)	0.478	(0.018)	0.496	(0.019)	0.360	(0.018)	-0.107	(0.008)
	2 (BED)	0.571	(0.023)	0.953	(0.039)	1.038	(0.050)	0.718	(0.032)	-0.162	(0.013)
	3 (MOB)	0.489	(0.016)	0.794	(0.030)	0.810	(0.031)	0.663	(0.029)	-2.088	(0.339)
	4 (DRS)	0.569	(0.029)	0.839	(0.034)	0.845	(0.038)	0.675	(0.030)	-0.134	(0.009)
	5 (BTH)	0.401	(0.014)	0.624	(0.021)	0.595	(0.022)	0.594	(0.023)	-2.102	(0.351)
	6 (TLT)	0.450	(0.018)	0.770	(0.029)	0.842	(0.035)	0.702	(0.030)	-0.118	(0.008)
$Age_{1/2}$	1 (EAT)	103.493	(0.429)	92.291	(0.222)	85.119	(0.170)	78.610	(0.225)	65.167	(0.731)
	2 (BED)	97.348	(0.190)	87.315	(0.123)	81.092	(0.099)	73.717	(0.125)	89.701	(0.722)
	3 (MOB)	95.422	(0.165)	86.036	(0.119)	79.990	(0.101)	72.287	(0.125)	95.050	(0.240)
	4 (DRS)	99.291	(0.238)	88.516	(0.137)	81.955	(0.117)	74.622	(0.132)	79.339	(0.544)
	5 (BTH)	95.340	(0.178)	85.406	(0.120)	79.003	(0.112)	71.454	(0.119)	94.832	(0.274)
	6 (TLT)	99.021	(0.259)	87.703	(0.136)	81.335	(0.106)	74.130	(0.126)	83.221	(0.678)

Table 3.5: Posterior means for parameters of interest for basic model with $K = 5$ extreme profiles. Numbers between parenthesis are posterior standard deviations.

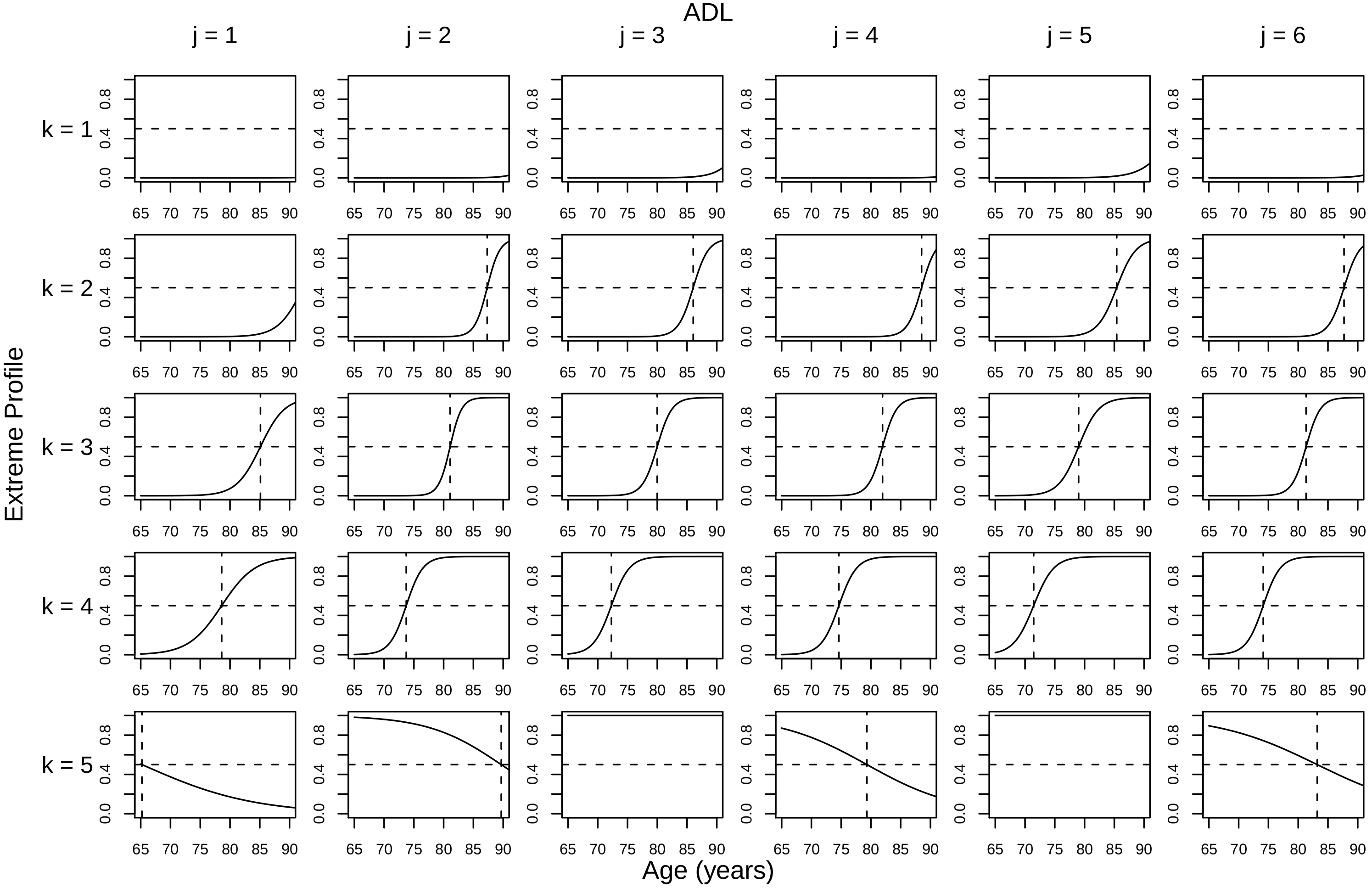


Figure 3.5: Extreme trajectories of disability over time for each ADL and extreme profile ($K = 5$) Dashed lines indicate the point where the probability of acquiring a disability in ADL j reaches 50%.

			Extreme Profiles			
Age Range	Pattern	Observed	$K = 2$	$K = 3$	$K = 4$	$K = 5$
$65 - 70$	000000	21361	19922	20635	21042	21249
	111111	295	0	11	45	175
	010000	292	554	301	243	288
	111110	192	5	44	108	260
	000100	167	447	220	157	172
	010110	117	62	103	95	68
	010100	103	180	142	92	72
	000110	89	101	81	59	64
$70 - 79$	000000	34508	32148	33340	33550	33773
	010000	1143	1217	1386	1222	1232
	111111	1087	156	403	378	429
	111110	692	407	733	738	789
	000100	584	779	734	670	624
	010100	412	413	332	427	411
	010110	375	345	252	286	282
	110110	335	358	356	365	384
$79 - 80$	000000	15536	13382	13939	14354	14600
	111111	1734	818	821	997	1159
	010000	1178	1845	1349	1385	1251
	111110	1018	920	995	1065	911
	000100	562	1251	881	839	725
	010100	536	413	534	503	550
	010110	529	256	373	355	431
	110110	461	349	441	448	465
$90 <$	000000	1311	369	736	904	907
	111111	1025	541	793	858	860
	111110	529	303	351	450	463
	010110	196	161	126	115	130
	010000	193	287	338	315	288
	010100	189	246	211	179	190
	110110	158	145	114	114	128
	000100	138	254	281	262	235
		X^2	141970	15343	7516	4638

Table 3.6: Observed and mean predictive posterior aggregated counts for the 10 most populated response patterns by age range, for basic model with $K = 2, 3, 4, 5$. Within each age group the response patterns are sorted decreasingly according to observed frequencies. For each model the X^2 statistic has been computed from the complete $2^6 \times 4$ contingency table.

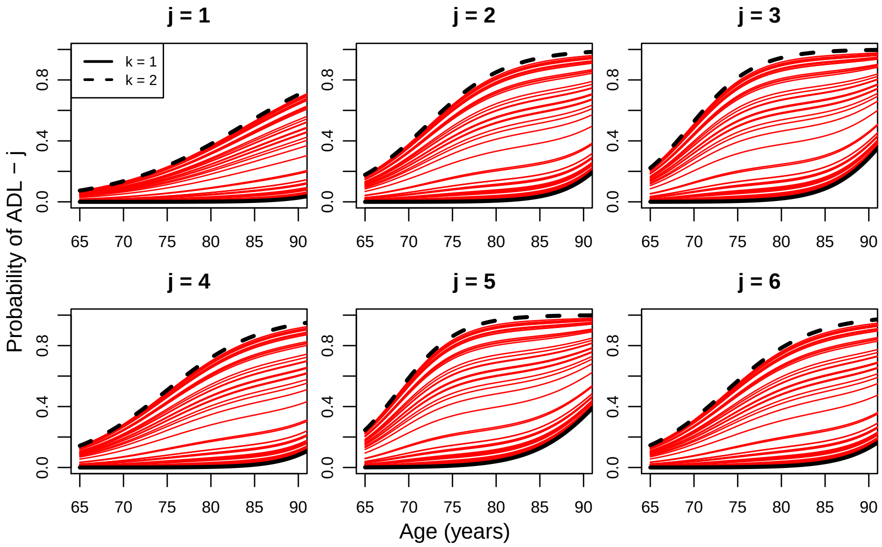


Figure 3.6: Individual-level mixture of trajectories for model with $K = 2$ extreme profiles for each ADL. Extreme trajectories are represented with thick lines and and a random sample of 100 individual posterior trajectory curves are plotted using thin lines

that the variant is exactly the Basic Model, but with extreme trajectories, $\lambda_{jk}(Age)$, modeled using step functions.

For the trajectories, I have defined the partition of the age range for constructing the step function with breakpoints $c_1 = 68, c_2 = 71, c_3 = 74, c_4 = 77, c_5 = 80, c_6 = 83$ and $c_7 = 86$. This particular choice of break points defines trajectories over time with 8 steps per combination of ADL and extreme profile, and therefore $8 \times J \times K$ parameters to describe them.

For the trajectory parameters I have selected Beta$(1, 1)$ priors, as discussed in Chapter 2. For the parameters α, I have used the same priors as in the Basic Model with continuous trajectories.

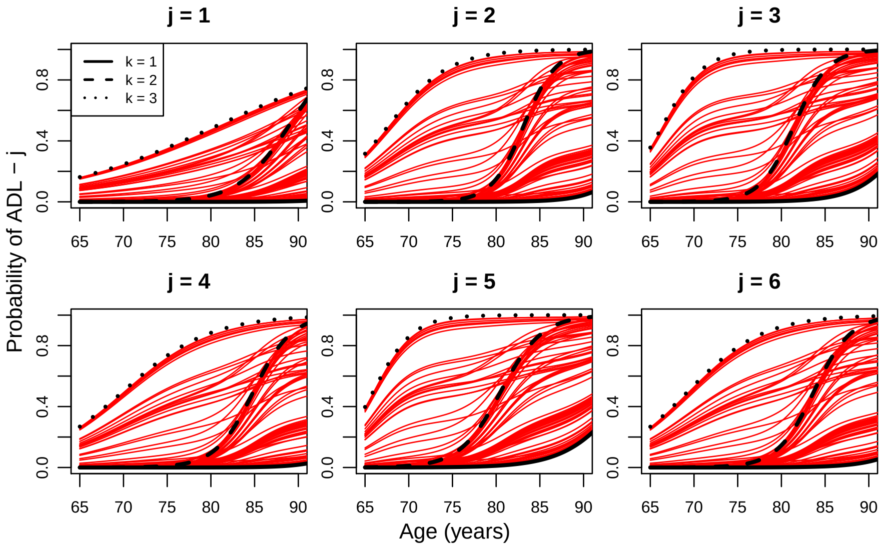


Figure 3.7: Individual-level mixture of trajectories for model with $K = 3$ extreme profiles for each ADL. Extreme trajectories are represented with thick lines and and a random sample of 100 individual posterior trajectory curves are plotted using thin lines

Applying the algorithm, we obtain posterior estimates for α_0 and ξ (Table 3.7) that are almost identical to the ones obtained through the application of the basic model with continuous trajectories in Section 3.2 (cf. Tables 3.2, 3.3, 3.4 and 3.5).

Figures 3.9, 3.10 and 3.11 present the estimated posterior extreme trajectories for models with $K = 3, 4$ and 5, overlaid with the continuous trajectories estimated from the Basic Model with continuous trajectories for comparison. The comparison is quite instructive. In all cases the discrete and continuous trajectories for the most predominant extreme profiles ($k = 1, 2$) are essentially the same. Extreme profile $k = 3$ in models with $K = 3$ (Figure 3.9, last row) and extreme profile $k = 4$ in models with $K = 4$ (Figure 3.10, last row) show a divergence when it comes to late ages. This is a phenomenon that occurs at late ages in profiles where subjects have an increased probability of acquiring disabilities

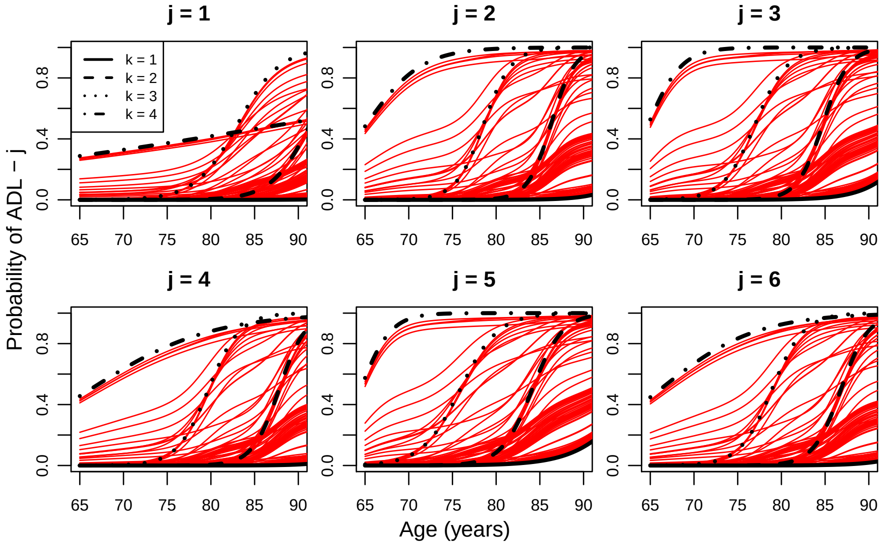


Figure 3.8: Individual-level mixture of trajectories for model with $K = 4$ extreme profiles for each ADL. Extreme trajectories are represented with thick lines and and a random sample of 100 individual posterior trajectory curves are plotted using thin lines

early in life. A likely explanation is that subjects for whom those characteristics are predominant, tend to die earlier than their peers that are closer to healthier profiles. I revisit this hypothesis in Section 7.

Looking at profile $k = 5$ for models with $K = 5$ extreme profiles (Figure 3.11) we see that all trajectories behave quite erratically. This situation mirrors that encountered in the application of the model with continuous trajectories in Section 3.2 (overlaid on the same figure), giving more evidence for the hypothesis of overfitting when $K > 4$.

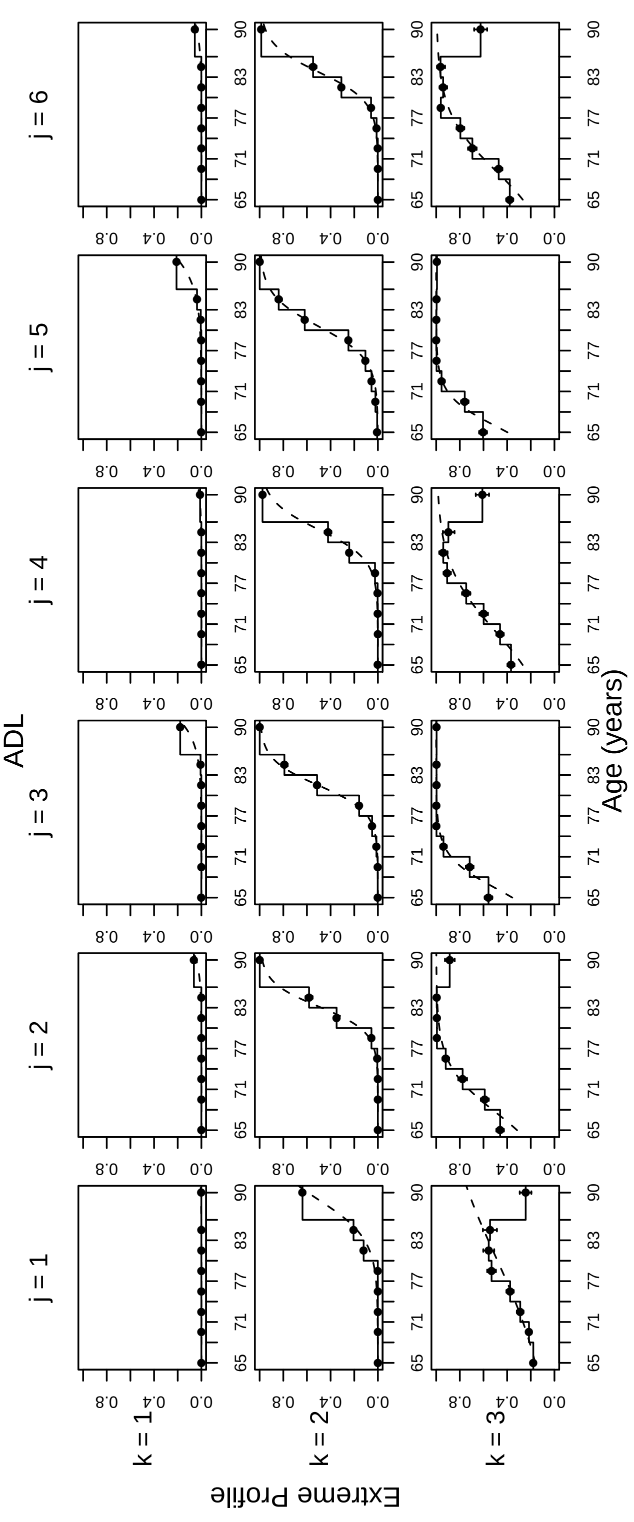


Figure 3.9: Extreme trajectories of disability over time for each ADL and extreme profile ($K = 3$) for models with piecewise constant and with continuous trajectories.

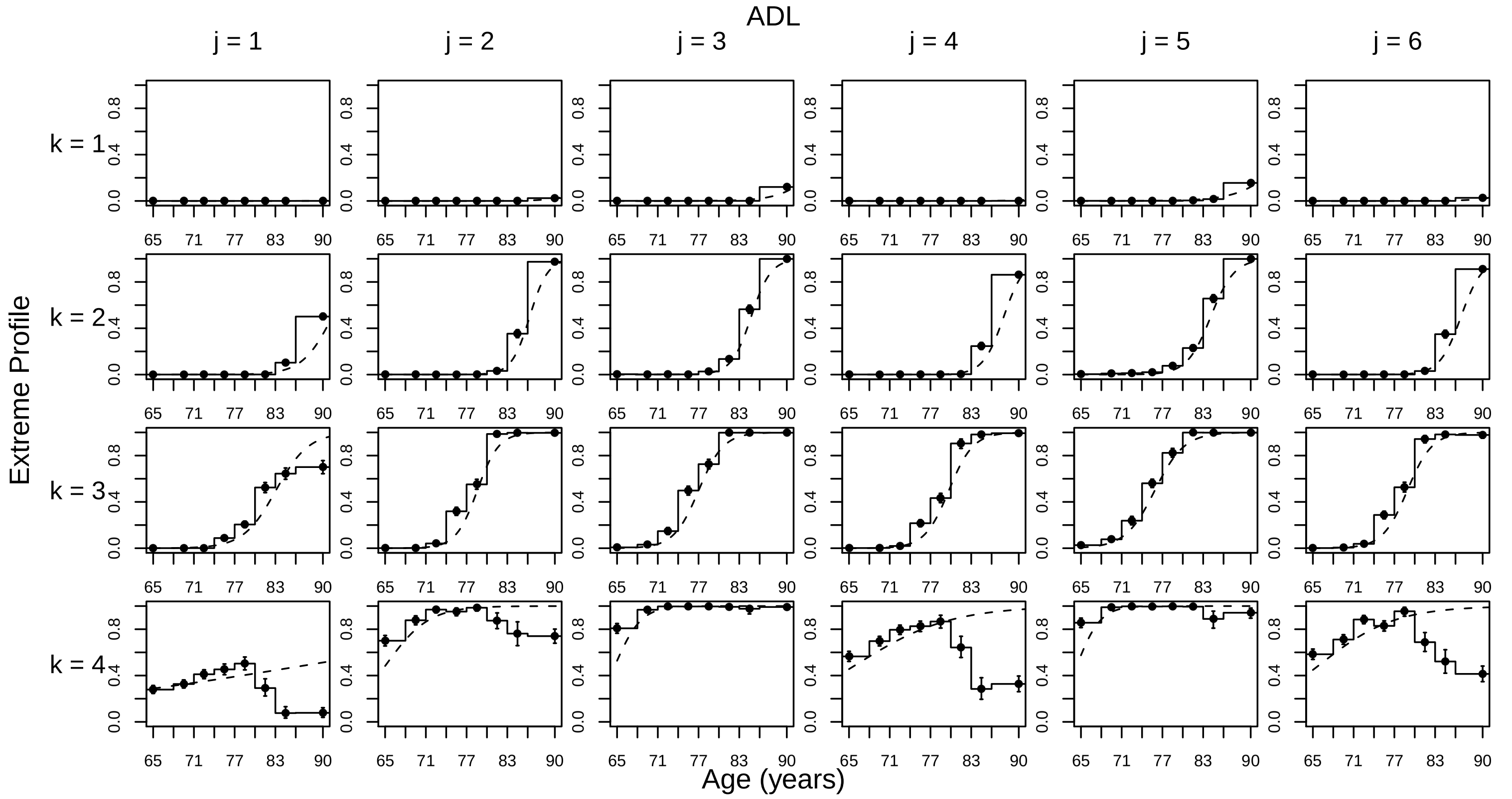


Figure 3.10: Extreme trajectories of disability over time for each ADL and extreme profile ($K = 4$) for models with piecewise constant and continuous trajectories.

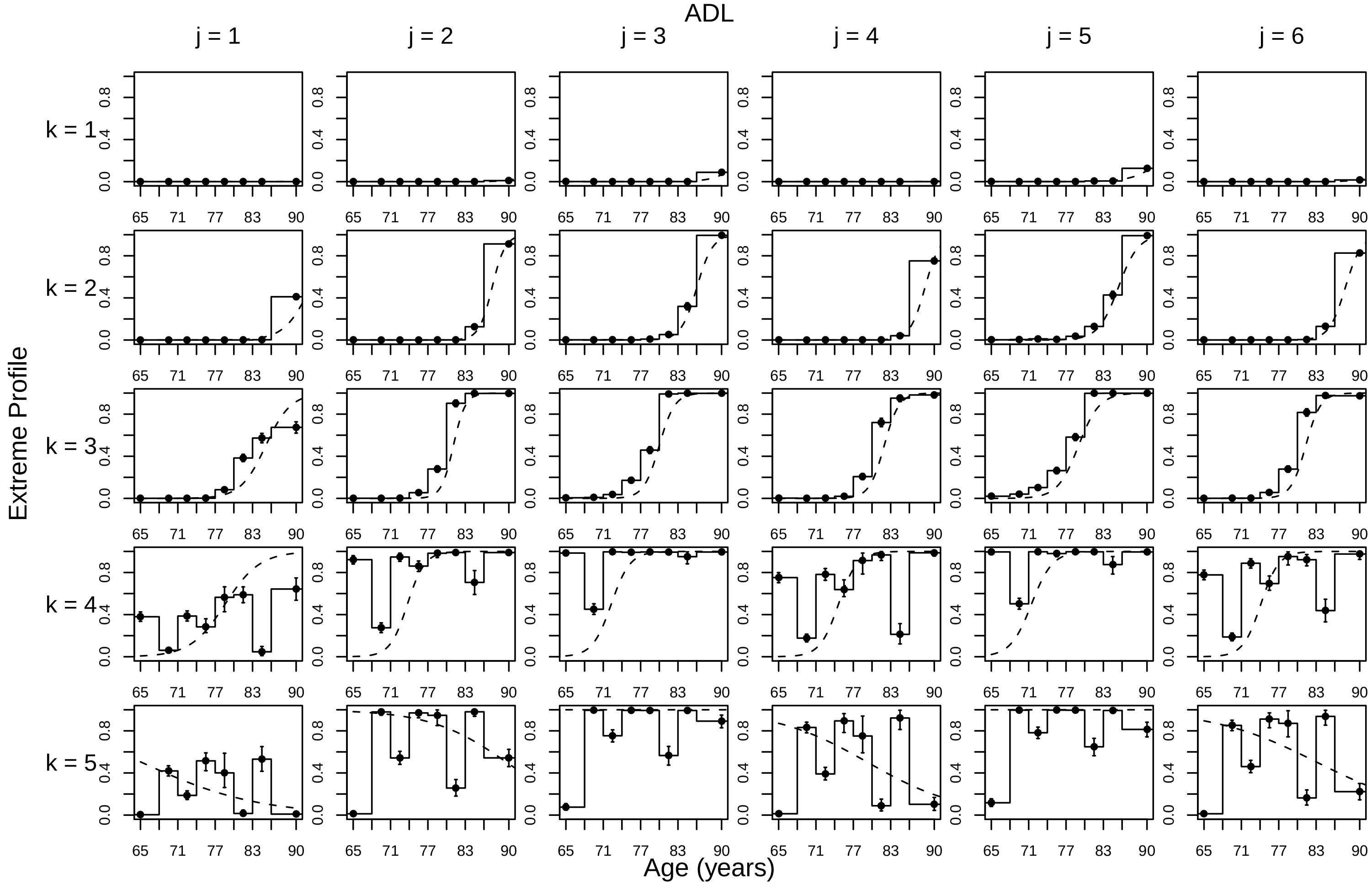


Figure 3.11: Extreme trajectories of disability over time for each ADL and extreme profile ($K = 5$) for models with piecewise constant and continuous trajectories.

Model	Parameter	Estimate [s.d.]
$K = 2$	α_0	0.325 [0.007]
	ξ	(0.820 [0.002], 0.180 [0.002])
$K = 3$	α_0	0.286 [0.006]
	ξ	(0.632 [0.004], 0.263 [0.003], 0.105 [0.002])
$K = 4$	α_0	0.269 [0.006]
	ξ	(0.558 [0.004], 0.256 [0.004], 0.117 [0.003], 0.069 [0.001])
$K = 5$	α_0	0.261 [0.006]
	ξ	(0.519 [0.005], 0.251 [0.005], 0.130 [0.003], 0.050 [0.001], 0.050 [0.002])

Table 3.7: Posterior means for parameters α_0 and ξ for basic model with discrete step trajectories for $K = 2, 3, 4, 5$ extreme profiles. Numbers between brackets are posterior standard deviations.

3.4 Discussion

The analyses I have presented in this chapter illustrate several desirable characteristics of the proposed method. They produce meaningful and easy to interpret summaries of the main trends in the population—the extreme trajectories—while at the same time producing estimates of the way in which the individuals are situated with respect to those extreme profiles, placing them as points over a continuous space defined by the simplex Δ_{K-1}. This approach fulfills two of the main goals outlined in the introduction of Chapter 2, namely interpretable summarization and handling of latent heterogeneity. Compared with hard classification techniques (cf. Connor 2006), my modelling approach allows me to keep the number of extreme trajectories small (and therefore easier to interpret), while at the same time providing enough flexibility to model complex trajectories that would otherwise not be well described by such a small number of extreme profiles. We must bear in mind that the method is in fact representing very large contingency tables ($2^{6\times 6}$ in the application), through the imposition of strong parametric assumptions and the use of covariates.

In Section 2.1, I emphasized that the choice of the trajectory functions is open to the modeler's decision, to fulfill the particular application needs. In this application I have applied both logistic s-shaped curves and step functions. The agreement between the estimates for the predominant profiles obtained with these two specifications provides evidence that the logistic trajectories are indeed a reasonable modeling decision. In particular, the inherent assumption of monotonicity is likely helping to make up for regions where the data are sparse to help in determining the trajectory, e.g., late ages in profile $k = 3$ in the model with $K = 3$ (Figure 3.9).

My results provide some interesting insights into the data. Most individuals end up placed close to a "healthy profile" (profile $k = 1$ in all applications of the model), where the trajectories reflect a basically healthy life until very late ages (90+). Then progressively, profiles with trajectories that reflect higher probabilities of disability at earlier ages are less and less predominant. In fact, if we permute the profiles, ordering them according to the expected value of the populational distribution of profile membership (component ξ_k for extreme profile k), the resulting ordering of trajectories will also show a progression from healthier trajectories to trajectories with earlier onset of disability.

The obtained sequence of disability manifestations [inside mobility→ toileting → dressing → bathing → getting in and out of bed → eating] within extreme profiles also provides further insight into disability patterns. This sequence, which was obtained exclusively from the data by sorting ADLs based on parameter $Age_{1/2,jk}$, resembles an intuitive ordering of tasks in terms of difficulty. It also seems universal, as it is exactly the same for each extreme profile within a model and for models with a different number of extreme profiles. These results demonstrate one of the positive features of the proposed models that is shared with other latent-trajectory approaches (c.f. Connor 2006).

One important characteristic of the model, that is well illustrated by the application, is that it handles all of the time dependence based on personal time (Age) opposed, to the introduction of either a "wave effect" (period effect) or cohort effect. Information about the waves is only used to determine the age of individuals, in order to place the responses into a comparable (personal) time scale. This point differentiates these methods from other longitudinal approaches, such as the one presented in White (2008), where the disability state transitions are estimated from one wave to another (basically a period effect). Chapter 4 will deal with the modeling and estimation of cohort effects.

3.A Appendix: Non-negative Priors for Trajectory Slope Parameters

In this appendix I modify the specification of the Basic model with continuous trajectories, presented in Section 2.2, constraining them to be nondecreasing. I have applied this modified version in Section 3.2 when discussing the possibility of excluding by design the cases of decreasing extreme trajectories.

A simple way of achieving this objective is to change the prior distribution of the slope parameter for each extreme trajectory, β_{jk}, selecting distributions with support on the non-negative real numbers, $\mathbb{R}^+$. A simple and convenient choice is the Gamma distribution:

$$\beta_{1jk} \overset{iid}{\sim} \text{Gamma}(a_1, b_1).$$

Analyzing the algorithm proposed in the previous Chapter, in Section 2.2.1, it is clear that the only step that is dependent on the prior distribution of β_{1jk} is Step 2, the sampling of the pair $(\beta_{0jk}, \beta_{1jk})$. This suggests that it is sufficient to modify that algorithm, replacing Step 2 by the following:

2′ **Sampling from $(\beta_{0jk}, \beta_{1jk})$:** Define $\Xi = \Xi_{tk}^{(ij)} = \{(i,j) : z_{ijt} = k\}$. The full conditional distribution of $(\beta_{0jk}, \beta_{1jk})$ is

$$
\begin{aligned}
&p\big((\beta_{0jk}, \beta_{1jk})|...\big)\\
&\quad \propto p(\beta_{0jk}, \beta_{1jk}) \times \prod_{\Xi} \frac{\exp\big[(\beta_0 + \beta_1 Age_{it})y_{ijt}\big]}{1 + \exp\big(\beta_0 + \beta_1 Age_{it}\big)}\\
&\quad \propto \exp\left[(a_1 - 1)\log\beta_1 - b_1\beta_1\right] \times \exp\left[-\frac{\beta_0^2}{2\sigma_0^2} + \frac{\beta_0\mu_0}{\sigma_0}\right] \times \exp\left[\beta_0 \sum_{\Xi} y_{ijt} + \sum_{\Xi} y_{ijt} \cdot Age_{it}\right]\\
&\quad \propto \exp\left[-\frac{\beta_0^2}{2\sigma_0^2} + \beta_0\left(\frac{\mu_0}{\sigma_0} + \sum_{\Xi} y_{ijt}\right) + \beta_1\left(-b_1 + \sum_{\Xi} Age_{it} \cdot y_{ijt}\right) + (a_1 - 1)\log\beta_1\right]
\end{aligned}
$$

Since this distribution is non-standard, we can apply the following Metropolis-Hastings step:

(a) *Proposal Step*: Sample the proposal values

$$
\begin{aligned}
\beta_0^* &\sim N\left(\beta_0, \sigma_0^{*2}\right)\\
\beta_1^* &\sim \text{lognormal}\left(\log\beta_1, \sigma^{*2}\right)
\end{aligned}
$$

(b) *Rejection Step*: Different from the original algorithm, this proposal distribution is not symmetric anymore, then we have to include the Hastings correction term

$$
\begin{aligned}
r =& \frac{p(\beta_0^*, \beta_1^*|...)}{p(\beta_0^*, \beta_1^*|...)} \frac{q(\beta_1|\beta_1^*)}{q(\beta_1^*|\beta_1)} \\
=& \prod_{\Xi} \left[\frac{1 + \exp\left[(\beta_0 + \beta_1 Age_{it})\, y_{ijt}\right]}{1 + \exp\left[(\beta_0^* + \beta_1^* Age_{it})\, y_{ijt}\right]} \right] \\
& \times \exp\left[-\frac{\beta_0^{*2} - \beta_0^2}{2\sigma_0^2} + (\beta_0^* - \beta_0) \left(\frac{\mu_0}{\sigma_0} + \sum_{\Xi} y_{ijt} \right) \right] \\
& \times \exp\left[(a_1 - 1)\big(\log \beta_1^* - \log \beta_1\big) + \big(\beta_1^* - \beta_1\big) \left(-b_1 + \sum_{\Xi} Age_{it} \cdot y_{ijt} \right) \right] \frac{\beta_1^*}{\beta_1} \\
=& \prod_{\Xi} \left[\frac{1 + \exp\left[(\beta_0 + \beta_1 Age_{it})\, y_{ijt}\right]}{1 + \exp\left[(\beta_0^* + \beta_1^* Age_{it})\, y_{ijt}\right]} \right] \\
& \times \exp\left[-\frac{\beta_0^{*2} - \beta_0^2}{2\sigma_0^2} + (\beta_0^* - \beta_0) \left(\frac{\mu_0}{\sigma_0} + \sum_{\Xi} y_{ijt} \right) \right] \\
& \times \exp\left[a_1\big(\log \beta_1^* - \log \beta_1\big) + \big(\beta_1^* - \beta_1\big) \left(-b_1 + \sum_{\Xi} Age_{it} \cdot y_{ijt} \right) \right]
\end{aligned}
$$

It is worth noting that this expression is exactly the same as Equation (2.12), except for the last line. Thus, the implementation of this algorithm should be straightforward, just slightly modifying the original sampler from Section 2.2.1.

Chapter 4

Modeling Generational Differences: Grouped Data Model

4.1 Introduction

In this chapter I propose an extension of the Basic Trajectory GoM model introduced in Chapter 2 that introduces birth cohort effects as part of the factors that explain the observed patterns of disability. The introduction of these effects is directed at answering questions about changes in the ways of aging across different generations, e.g. to provide answers to questions like "are younger generations acquiring disabilities differently than older ones?"

Up to this point in the thesis, the Basic GoM Trajectory Model proposed in Chapter 2 attributes all variation over time to the natural individual progression of aging. The observable differences in the prevalence of disability between different epochs are then explained by the aggregation of individuals that are at distinct points of their life-trajectories at each epoch. Although this formulation explicitly assumes a great deal of individual-level heterogeneity, modeled through the introduction of a partial-membership structure, a key assumption is that this heterogeneity can be fully explained by a common population-level underlying probability distribution, regardless of other factors.

When formulating assumptions to model changes over time we can basically distinguish three different ways of introducing that variability. We can assume that it is an age-related or “time in the system” phenomenon. This was the approach that I took when I constructed the Basic GoM trajectory model. Another perspective can be obtained by the introduction of a “period effect”: an external event or process over absolute time that affects all individuals simultaneously. Finally we can analyze it as a cohort effect. In this context by “cohort” we understand the same as with the term “generation”: a group of individuals that were born within the same period (Ryder, 1985).

There is a fundamental tension between the modeling of age, cohort and period effects that arises from the fact that they are linearly dependent. In fact, a person’s age is defined to be the time between the assessment date (period) and their date of birth: $Age_{it} = Date_t - DOB_i$. This can create an identification problem if the corresponding effects are linearly related, and appropriate identifiability constraints are not imposed (Fienberg and Mason, 1985). Although this is a fundamental problem that, when present, has to be addressed explicitly, the models I present in this section do not consider any period effects. Thus, I will not consider this issue any further.

4.2 Cohorts as covariates

Birth cohorts are defined by date intervals or grouping of possible date of births. Then, after having defined those groups, an individual's generation will be completely determined by the date of birth (DOB). If we want to assess the effect of the individual's generational group in their way of acquiring disabilities, a direct way of doing it is to extend our models to include the DOB as a covariate.

There are basically two alternatives when it comes to covariate placement in a model like the GoM trajectory model. We could place the covariates affecting directly the response probabilities at the level of extreme profiles. This is the approach I have already taken with the *Age* covariate, to form the extreme trajectories, $\lambda_{jk}(Age)$. The other alternative is to somehow make the membership structure dependent on the covariates.

In this section I explore the second alternative, making the membership structure dependent on the covariates. The reasons for this choice are mostly interpretative. In the first place, I have regarded each individual's classification, expressed by the membership vector, g_i, as a static characteristic of each individual. Making it dependent on other static individual characteristics can be viewed as a way of using additional information to improve the classification expressed by the individual membership.

The second reason is that such an approach will allow us to keep the extreme trajectories the same for the whole population. This provides us with a stable frame of reference to compare individuals with each other and *as part of generational groups*. The population can then be stratified by generations and we can compare the distribution of the membership vectors across different strata. Finally, we can read differences in the ways of aging as differences in the underlying distribution of membership, that can be interpreted using the common frame of reference provided by the extreme trajectories.

4.3 Extending the Basic Model to Handle Generational Groups

To implement the approach just outlined, to make the membership vectors dependent on DOB, we have to decide how will the DOB covariate will be related to the membership structure. Since we are interested in comparing the underlying distribution of membership, conditional on generational group, a direct way of enabling such comparisons is to keep the individual level structure proposed in the Basic Model, but replace the common distribution of membership vectors by a family indexed by a (possibly stochastic) function the covariate, DOB:

$$\begin{aligned} \Pr\left(Y_{ijt}=y_{ijt}\left|g_i,X_i,\theta\right.\right) &= \sum_{k=1}^{K} g_{ik} f_{\theta_{j|k}}(y_{ijt}|Age_{it}), \\ g_i|DOB_i &\sim G_{\alpha(DOB_i)}. \end{aligned}$$

I will keep the specification of the distribution of membership vectors introduced for the Basic model, using the Dirichlet distribution indexed by the parameter α, by specifying the family $G_{\alpha(DOB)} = \text{Dirichlet}(\alpha_{(DOB)})$. Then the indexing parameter will be the function $\alpha : \mathbb{R} \rightarrow \mathbb{R}^{+K}$ defined by

$$\alpha(DOB) = \big(\alpha_1(DOB), \alpha_2(DOB), ..., \alpha_K(DOB)\big). \tag{4.1}$$

A simple, yet reasonably flexible, way of specifying this relationship is to define a partition of the population into intervals, defined by ranges of dates of birth, and assume the membership distribution indexing parameter, α, constant for each interval on the partition. Let $\Gamma = \{\gamma_1, \gamma_2, ..., \gamma_C\}$ be a finite partition (contiguous non overlapping intervals) of the range of possible dates of birth, similar to the partition of ages defined in Section 2.3 for the construction of the discrete extreme trajectories. This partition assigns a particular generational group (or interval of dates of birth) to each individual in the sample. Then, within each subset assume a homogeneous distribution of membership vectors, by assuming a constant indexing parameter for G_α, $\alpha_{(\gamma)*}$. Now we can define the generic k-th component of the function $\alpha(DOB)$ by

$$\alpha_k(DOB; \alpha_{(*)k}) = \prod_{\gamma \in \Gamma} \left[\alpha_{(\gamma)k}\right]^{I(DOB \in \gamma)}, \tag{4.2}$$

where $\alpha_{(\gamma)k} > 0$. The argument after the semi-colon in the notation, $\alpha_k(DOB; \alpha_{(*)k})$, is just a reminder of the fact that we are introducing a number of new parameters: as many K-dimensional vectors as generational groups are defined in Γ.

This specification is very similar to the one used in the Basic model, just taking a less constrained distribution for the membership vectors. In more abstract terms we have that DOB is just an instance of an individual-level time-invariant covariate, of the same type as those introduced in Chapter 1, Section 1.2, under the generic symbol X_i^f. Assuming that these covariates are independent and identically distributed samples from a common population-level distribution, F_X, it is clear that the proposed extension is just replacing the probability measure associated with the individual membership scores, G, by the mixture

$$G(\cdot) = \int G(\cdot|x) F_X(dx), \tag{4.3}$$

which in this specific case is just the discrete mixture:

$$G(\cdot) = \sum_{\gamma \in \Gamma} p(\gamma) G_{\alpha_{(\gamma)}}(\cdot). \tag{4.4}$$

Since we are actually interested in the dependency between the distribution and the covariate DOB, I just leave the specification of the distribution G conditional on DOB. Extending the Basic model to handle generational information thus consists of replacing the population level distribution of membership vectors, $p(g_i|\alpha)$ by the conditional version, $p\big(g_i|\alpha(DOB_i; \alpha_{(*)*})\big)$.

I consider that all of the newly introduced parameters come from a common distribution, in the same way that I have previously assumed that the parameter α was a single draw from a hyperdistribution in the Basic Model.

To this end, define

$$\alpha_{(\gamma)0} = \sum_{k=1}^{K} \alpha_{(\gamma)k} \quad \text{and} \quad \xi_{(\gamma)k} = \frac{\alpha_{(\gamma)k}}{\alpha_{(\gamma)0}},$$

so that we can again use the priors

$$\alpha_{(\gamma)0} \overset{iid}{\sim} \text{Gamma}(\tau, \eta) \quad \text{and} \quad \xi_{(\gamma)*} \overset{iid}{\sim} \text{Dirichlet}(\mathbf{1}_K),$$

just as with the Basic Model.

4.4 Estimation

The posterior distribution of parameters conditional on the data and covariates is quite similar to the basic model:

$$p(\alpha,\beta,g|y,z,Age,DOB) \propto p(\beta)p(\alpha)\underbrace{\left(\prod_{i=1}^{N} p\left(g_{i*}|\alpha(DOB_i;\alpha_{(*)*})\right)\right)}_{(*)}$$
$$\times \prod_{i=1}^{N}\prod_{j=1}^{J}\prod_{t\in M_i}\prod_{k=1}^{K}\left(g_{ik}\frac{\exp(y_{ijt}\beta_{0jk}+y_{ijt}\beta_{1jk}Age_{it})}{1+\exp(\beta_{0jk}+\beta_{1jk}Age_{it})}\right)^{I(z_{ijt}=k)},$$

under the same considerations as in Chapter 2.2.1.

If we replace the expression $\alpha(DOB)$ by the partition-wise constant specification given by Equation (4.2), the expression labeled $(*)$ becomes

$$\begin{aligned}(*) &= \prod_{i=1}^{N} p\big(g_i|\alpha(DOB_i)\big)\\ &= \prod_{\gamma\in\ \Gamma}\prod_{i=1}^{N}\left[p\big(g_i|\alpha_{(\gamma)*}\big)\right]^{I(DOB_i\in\gamma)} \qquad (4.5)\end{aligned}$$

From the structure of this distribution, it is clear that most of the steps of the MCMC algorithm developed for the Basic Model can be reused and the rest adapted. The samplers that have to be adapted are those corresponding to α and g_i.

Let $\gamma_i \in \Gamma$ be the unique interval from the partition such that $DOB_i \in \gamma_i$.

1. **Sampling from** g_i. The full conditional distribution of g_i is

$$p(g_i|...) \propto p\left(g_i\left|\alpha(DOB_i)\right.\right) \times \prod_{k=1}^{K} g_{ik}^{\sum_{j,t} I(z_{ijt}=k)}.$$

Using the expression from Equation (4.5), we get

$$\begin{aligned}
p(g_i|...) &\propto \prod_{\gamma \in \ \Gamma} \left[p\big(g_i|\alpha_{(\gamma)*}\big)\right]^{I(DOB_i \in \gamma)} \times \prod_{k=1}^{K} g_{ik}^{\sum_{j,t} I(z_{ijt}=k)} \\
&= p\big(g_i|\alpha_{(\gamma_i)*}\big) \times \prod_{k=1}^{K} g_{ik}^{\sum_{j,t} I(z_{ijt}=k)} \\
&= \text{Dirichlet}(g_i|\alpha_{(\gamma_i)*}) \times \prod_{k=1}^{K} g_{ik}^{\sum_{j,t} I(z_{ijt}=k)} \\
&= \text{Dirichlet}(g_i|\alpha_{(\gamma_i)*}) \times \prod_{k=1}^{K} g_{ik}^{\sum_{j,t} I(z_{ijt}=k)} \\
&= \text{Dirichlet}\left(\alpha_{(\gamma_i)1} + \sum_{j,t} I(z_{ijt}=1), ..., \alpha_{(\gamma_i)K} + \sum_{j,t} I(z_{ijt}=K)\right).
\end{aligned}$$

2. **Sampling from** $\alpha_{(\gamma)*}$ Define $\Xi_\gamma = \{i : DOB_i \in \gamma\}$. The full conditional distribution of $\alpha_{(\gamma)*}$ is

$$\begin{aligned}
p(\alpha_{(\gamma)*}|...) &\propto p(\alpha_{(\gamma)*}) \times \prod_{i=1}^{N} \left[p\big(g_i|\alpha_{(\gamma)*}\big)\right]^{I(DOB_i \in \gamma)} \\
&\propto \text{Gamma}(\alpha_{(\gamma)0}|\tau, \eta) \times \text{Dirichlet}(\xi_{(\gamma)}|\mathbf{1}_K) \times \prod_{i=1}^{N} \text{Dirichlet}(g_i|\alpha_{(\gamma)*})^{I(DOB_I \in \gamma)} \\
&\propto \alpha_{(\gamma)0}^{\tau-1} \exp[-\alpha_{(\gamma)0}\eta] \times \prod_{i=1}^{N} \left[\frac{\Gamma\left(\alpha_{(\gamma)0}\right)}{\prod_{k=1}^{K} \Gamma(\alpha_{(\gamma)k})} \prod_{k=1}^{K} g_{ik}^{\alpha_{(\gamma)k}}\right]^{I(DOB_I \in \gamma)} \\
&\propto \alpha_{(\gamma)0}^{\tau-1} \exp[-\alpha_{(\gamma)0}\eta] \times \left[\frac{\Gamma\left(\alpha_{(\gamma)0}\right)}{\prod_{k=1}^{K} \Gamma(\alpha_{(\gamma)k})}\right]^{\#(\Xi_\gamma)} \prod_{k=1}^{K} \left[\prod_{\Xi_\gamma} g_{ik}\right]^{\alpha_{(\gamma)k}},
\end{aligned}$$

where $\#(\Xi_\gamma)$ is the number of elements in the set Ξ_γ or, equivalently, the number of individuals with Date of Birth within the interval γ.

This expression is similar to the corresponding one in the Basic Model (Equation 2.13): we are only replacing all the iterations over the whole sample (i.e. $i = 1...N$) to iterations over the individuals whose DOBs fall within the interval γ. This suggests adapting the procedure in the algorithm by just replacing r, in step 4, Chapter 2 by

$$r = \min\left\{1, \exp\left[-\tau(\alpha_0^* - \alpha_{(\gamma)0})\right] \left(\prod_{k=1}^{K} \frac{\alpha_k^*}{\alpha_{(\gamma)k}}\right) \left(\frac{\alpha_0^*}{\alpha_{(\gamma)0}}\right)^{\tau-1} \right.$$
$$\left. \times \left[\frac{\Gamma(\alpha_0^*)}{\Gamma(\alpha_{(\gamma)0})} \prod_{k=1}^{K} \frac{\Gamma(\alpha_{(\gamma)k})}{\Gamma(\alpha_k^*)}\right]^{\#(\Xi_\gamma)} \prod_{k=1}^{K} \left(\prod_{i\in\Xi_\gamma} g_{ik}\right)^{\alpha_k^* - \alpha_{(\gamma)k}} \right\}$$

and update the chain, from step m to step $m+1$ according to the rule

$$\alpha_{(\gamma)*}^{(m+1)} = \begin{cases} \alpha^* & \text{with probability} \quad r. \\ \alpha_{(\gamma)*}^{(m)} & \text{with probability} \quad 1-r. \end{cases}$$

4.5 Further Remarks

In this chapter, I have proposed a simple extension of the Basic GOM Trajectory model that can accommodate information about birth cohorts. Keeping the extreme trajectories the same for the whole population enables us to use them as a fixed references for comparison between different cohorts. Additionally, it enables us to use data from the entire sample to reconstruct them, increasing the quality of the estimates.

Although the proposed extension was explicitly developed to assess differences between different generational groups, the method presents a generic way of introducing time-invariant covariates by means of stratifying the population into disjoint classes. It can be used unchanged to handle any sort of strata whose effect could be modeled as a particular tendency in the membership structure of its individuals. A natural use of this idea is to model the effect attributable to other time invariant covariates, like sex or even sampling-related strata.

This idea can also be extended in a more sophisticated way by specifying other parametric dependence relationships between the membership distribution and the covariates, either by specifying the parameters of the Dirichlet distribution as another parametric function of the covariates,$\alpha(X)$, or by replacing the prior distribution of the membership scores, G, by another distribution conditional on the covariates, like the multilevel-logistic prior proposed in Bertolet (2008) and Blei and Lafferty (2007).

This model is deliberately attributing all the systematic time variation in the disability responses to age and cohort. While this specification produces a richer and more versatile model than just ignoring cohort effects, as in Chapter 2, relevant period effects could be ignored and fudged into the other effects. In the study of disability, effects such as medical breakthroughs or changes in public policy are obvious reasons to believe that period effects might be relevant.

Chapter 5

Applying the Grouped Model to the NLTCS

In this chapter, I apply the extension of the GoM trajectory model proposed in Chapter 4 to the NLTCS data. To this effect I define an arbitrary partition of the range of dates of birth present in the NLTCS sample (cohorts) and use the Grouped GoM trajectory model to estimate the underlying distribution of membership vectors, conditional on the generational groups. Finally, I use the estimates of the parameters of the population-level distribution, conditional generational on membership, to draw conclusions about those generations.

Cohort	DOB	1982 ($t = 1$)	1984 ($t = 2$)	1989 ($t = 3$)	1994 ($t = 4$)	1999 ($t = 5$)	2004 (*) ($t = 6$)
1	1	6329	6025	1347	1397	617	70
2	2	7631	7082	3452	3335	1753	575
3	3	3696	7839	2627	5102	3679	2010
4	4	1	463	2410	4581	4724	3505
5	5	0	0	0	2478	6403	4251

(*) Only individuals present in 1999

Table 5.1: Definition of cohorts and number of measurements at each wave. Cohort 5 only have measurements in the last three waves of the NLTCS.

5.1 Preliminaries

The data used in this chapter is the same that was used in Section 3 to generate the estimates for the Basic GoM model and thus the reader should refer to Chapter 3 for details.

In addition, I have defined five cohorts or generational groups, partitioning the ranges of possible dates of birth according to the intervals defined in the first two columns of Table 5.1. I selected these intervals so that they group approximately the same number of individuals from the sample.

Table 5.1 also shows the number of measurements at each wave, for each cohort. A salient feature of this arrangement is that individuals from the youngest cohort (cohort 5—born after 1926) have measurements only in the last three waves due to age eligibility, as its oldest members turned 65 after 1991.

It is also worth noting that by their definition, neither the oldest (cohort 1—born before 1906) nor the youngest (cohort 5—born after 1926) cohorts spans the whole range of

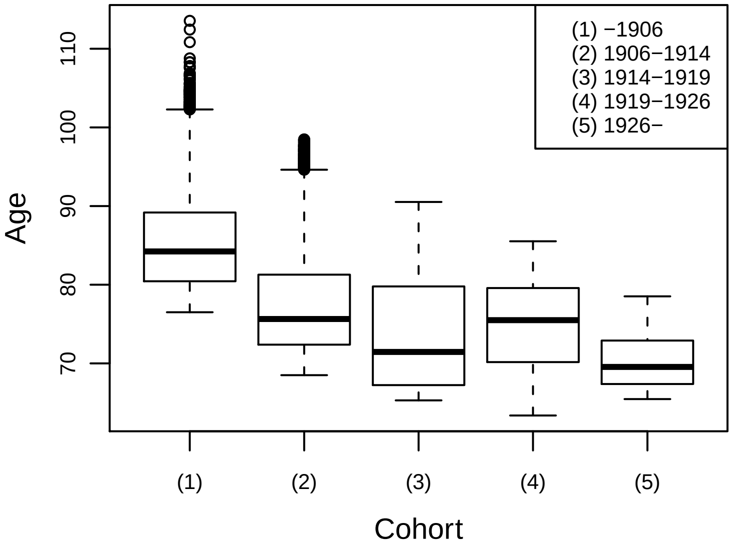


Figure 5.1: Distribution of age coverage for all defined generational groups.

relevant DOBs in the NLTCS. In fact, the oldest individual in cohort 5 can be at the most 78 years old in 2004, while the youngest individual from cohort 1 cannot be younger than 76 years old in 1982. Figure 5.5 shows the distribution of pooled ages from the NLTCS sample, at every wave, stratified by generational group.

I selected the prior distributions to match those of the Basic model, in Chapter 3, in part because the same considerations discussed there apply in this case, and to allow comparisons, as I am considering this model as an extension of the original basic GoM trajectory model.

The execution of the algorithm is very similar to the execution of the algorithm corresponding to the basic model and the same considerations in terms of proposal distributions for the common parameters apply. While it makes sense to consider different tuning parameters for the proposal distributions of each cohort's membership distribution parameter,

$\alpha_{(\gamma)*}$, in practice it is enough to tune the sampler using a common proposal distribution.

5.2 Results

The main objective of the analysis with this model is to compare the underlying distribution of the membership vectors conditional on generational groups. We can do this by directly comparing the parameters of these distributions for each generational group $\gamma \in \Gamma$, $\alpha_{(\gamma)*}$ and interpret them with respect to the common extreme trajectories, defined by the parameters β.

The common extreme profile parameters, β, are very similar to those computed using the basic model, and therefore we refer to the description in their corresponding section (Section 3.2 in Chapter 2). The corresponding posterior estimates under this model are in Tables 5.2 to 5.5 in the Appendix of this Chapter.

Figures 5.2 to 5.4 show the estimates (posterior means) of the components of the vector ξ for models with $K = 2, 3$ and 4 extreme profiles. For each generational group the sequence of values of ξ_k, reading from left to right are linked with lines. Read from left to right, these sequences indicate the evolution of the relative weight of the k-th component in each cohort, as we shift our attention from older to younger cohorts.

The most striking feature in these plots is the evident increasing monotonicity of the relative importance of the first component ($k = 1$) in each cohort, as we consider younger and younger cohorts, i.e. $\xi_{(1)1} < \xi_{(2)1} ... < \xi_{(5)1}$. This is especially clear in models with $K = 2$ and $K = 3$. In the model with $K = 4$, because of the high posterior dispersion it is not clear if the youngest generation actually follows this pattern. This trend is telling us that, as we consider newer cohorts, their members tend to be closer and closer to profile $k = 1$. This profile corresponds to the healthiest aging progression, with extremely low probability of acquiring disabilities until very advanced ages, as it can be observed, for instance, in Figure 3.3, in Chapter 3. Thus, as we consider younger generations, individuals tend to have healthier ways of aging, compared with their elders.

Further, we note that in these plots, the high posterior uncertainty of the estimates of the parameter ξ for the members of the last (youngest) generation. This situation starts to be apparent for $K = 3$ and for $K = 4$ is quite evident. This suggests that we should not take the apparent decrease from $\xi_{(4)1}$ to $\xi_{(5)1}$ too seriously, since the spread of the marginal posterior distribution of these parameters is too large. A likely explanation for this effect is the lack of data for ages past 78 years old in cohort 5, as we noted previously. I comment more on this in the next section.

To evaluate the fit of these models to the data, I have computed posterior predictive counts in the exact same way as I did for the Basic model. The results are presented in Table 5.6. When comparing these results to the ones obtained through the application of the Basic model (Table 3.6), we see that they are almost exactly the same. This could be expected given the similarity of the extreme profile estimates for both models and the fact that, although the Basic model is in fact more constrained than with the extended model, it still has a great degree of flexibility to handle the individual heterogeneity.

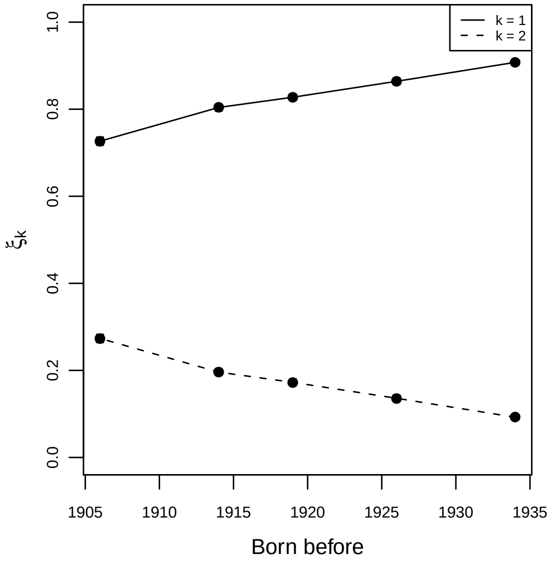


Figure 5.2: Evolution of the parameter vector ξ across different generations for model with $K = 2$ extreme profiles. The error bars show the 95% equal tail posterior credible intervals associated with the k-th component of the vector ξ.

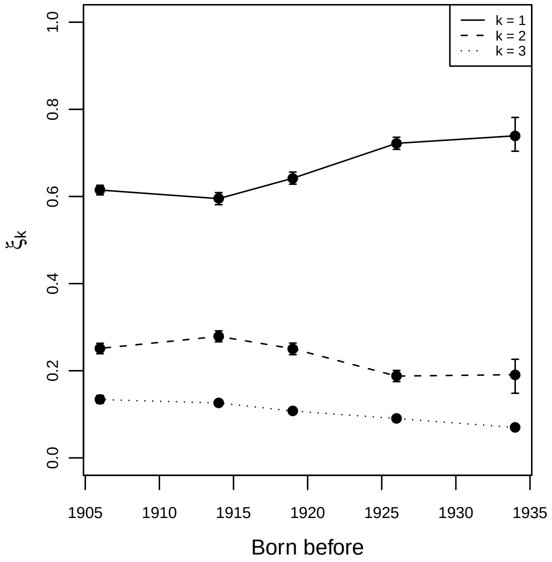


Figure 5.3: Evolution of the parameter vector ξ across different generations for model with $K = 3$ extreme profiles. The error bars show the 95% equal tail posterior credible intervals associated with the k-th component of the vector ξ.

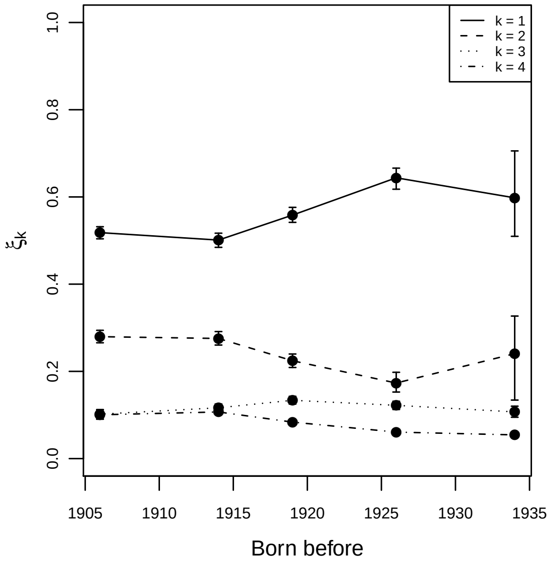


Figure 5.4: Evolution of the parameter vector ξ across different generations for model with $K = 4$ extreme profiles. The error bars show the 95% equal tail posterior credible intervals associated with the k-th component of the vector ξ.

5.3 Discussion

The application of the Grouped GoM trajectory model to the NLTCS data illustrates that it can be a powerful and easily interpretable tool in assessing individual differences attributable to cohort and in general, observed time-invariant characteristics of the individuals.

Because of the model formulation, requiring reading the generational differences with respect to population-level shared extreme profiles, we have to start by interpreting the extreme profiles themselves. After that we can read differences between cohorts as differences in the relative weight of each extreme profile in the population, directly from parameters $\xi_{(\gamma)k}$.

In terms of answering the motivating question "do younger generations acquire disabilities differently than older ones?", the answer provided by this exercise is "yes". Furthermore, there appears to exist a monotonic increase in the relative importance of a "healthy aging" or "late disability" profiles ($k = 1$), to the detriment of the other profiles, as we consider younger generations. This evidence points in the direction of a positive evolution in the ways of aging over time, where younger people are aging more "disability free" than people used to do in previous generations.

One limitation of this approach to assessing change is the need for long series of observations for each cohort. This problem is evident in the case of the last cohort, where we have only measurements until the age of 78. For models with $K > 2$ extreme profiles, the posterior distribution of parameters ξ corresponding to that cohort have a large spread, which in the case of $K = 4$ and $K = 5$ makes it impossible to assess the direction of the change with respect to the previous generation, or even if there was change at all.

The two cohorts where by design important age ranges are excluded are cohort 1, where people are older than 76 years old and; cohort 5, with members younger than 78 years old. Figure 5.5 and Figure 5.6 show the distribution of measured ages from these two generations compared with the three extreme trajectories of ADL 5, for models with $K = 3$ and $K = 4$ extreme profiles. It is interesting to note the regions in which the extreme trajectories are well separated. While for both models cohort 1 spans regions in which all extreme trajectories are well separated, the situation is different for cohort 5. For a model with $K = 3$ extreme profiles, separation between profile $k = 1$ and $k = 2$ can be visually assessed from age 70 (where $\lambda_2(70) - \lambda_1(70) \approx 0.012$), and by age 78 is quite evident ($\lambda_2(78) - \lambda_1(78) \approx 0.26$). However, for a model with $K = 4$ extreme profiles, those same differences are much smaller, i.e. $\lambda_2(70) - \lambda_1(70) < 10^{-4}$ and $\lambda_2(78) - \lambda_1(78) \approx 0.029$. This explains the high uncertainty of the posterior estimates of the membership in those

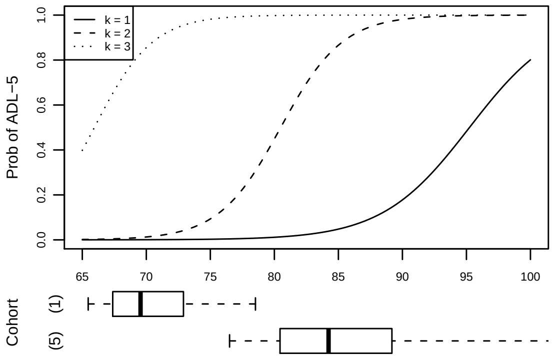


Figure 5.5: Comparison between the extreme trajectories for ADL-5 for a model with $K = 3$ extreme profiles and the distribution of age coverage for generations 1 (born before 1906) and 5 (born after 1926). Although by design none of these cohorts spans the whole interval of $65 - 90$ years, both include regions where the profiles are well separated.

profiles, for members of the last cohort, specially when $K = 4$.

The only possible "solution" to this situation within the proposed framework is to redefine the cohorts, so that all of them include regions where the difference between the profile curves is clear. However, at least for the model with $K = 3$ extreme profiles in this application to the NLTCS data, that uncertainty is not as big to justify the loss of cohort resolution.

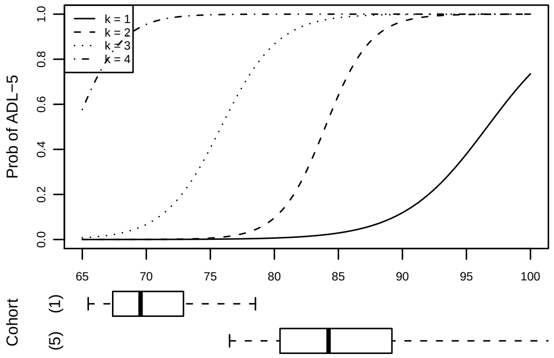


Figure 5.6: Comparison between the extreme trajectories for ADL-5 for a model with $K = 4$ extreme profiles and the distribution of age coverage for generations 1 (born before 1906) and 5 (born after 1926). Different from the case of $K = 3$, Generation 1 does not span a region where profiles $k = 1$ and $k = 2$ are well separated.

5.A Appendix: Posterior Estimates of Parameters

Parameter	Estimate
$\alpha_{0(5)}$	0.137 (0.013)
$\alpha_{0(4)}$	0.351 (0.016)
$\alpha_{0(3)}$	0.382 (0.015)
$\alpha_{0(2)}$	0.412 (0.015)
$\alpha_{0(1)}$	0.327 (0.011)

Parameter	Cohort	Estimate Extreme Profile-k (sd)			
		$k = 1$		$k = 2$	
ξ	5	0.908	(0.003)	0.092	(0.003)
	4	0.864	(0.004)	0.136	(0.004)
	3	0.827	(0.004)	0.173	(0.004)
	2	0.804	(0.004)	0.196	(0.004)
	1	0.727	(0.005)	0.273	(0.005)
Parameter	ADL(j)	Estimate Extreme Profile-k (sd)			
		$k = 1$		$k = 2$	
β_{0**}	1 (EAT)	-6.650	(0.132)	-0.518	(0.025)
	2 (BED)	-4.952	(0.068)	1.707	(0.052)
	3 (MOB)	-3.848	(0.044)	2.840	(0.087)
	4 (DRS)	-5.663	(0.091)	0.920	(0.038)
	5 (BTH)	-3.183	(0.033)	3.294	(0.101)
	6 (TLT)	-4.764	(0.062)	1.257	(0.044)
β_{1**}	1 (EAT)	0.293	(0.008)	0.130	(0.003)
	2 (BED)	0.312	(0.005)	0.211	(0.005)
	3 (MOB)	0.288	(0.004)	0.268	(0.008)
	4 (DRS)	0.310	(0.007)	0.177	(0.004)
	5 (BTH)	0.243	(0.003)	0.290	(0.009)
	6 (TLT)	0.276	(0.005)	0.196	(0.005)
$Age_{1/2}$	1 (EAT)	102.678	(0.329)	83.993	(0.239)
	2 (BED)	95.898	(0.141)	71.912	(0.148)
	3 (MOB)	93.382	(0.114)	69.398	(0.137)
	4 (DRS)	98.249	(0.188)	74.791	(0.156)
	5 (BTH)	93.110	(0.123)	68.624	(0.130)
	6 (TLT)	97.237	(0.174)	73.605	(0.146)

Table 5.2: Posterior means for parameters of interest for basic model with $K = 2$ extreme profiles. Numbers between parenthesis are posterior standard deviations.

Parameter	Estimate
$\alpha_{0(5)}$	0.253 (0.019)
$\alpha_{0(4)}$	0.194 (0.012)
$\alpha_{0(3)}$	0.224 (0.010)
$\alpha_{0(2)}$	0.310 (0.011)
$\alpha_{0(1)}$	0.291 (0.010)

Parameter	Cohort	Estimate Extreme Profile-k (sd)					
		$k = 1$		$k = 2$		$k = 3$	
ξ	5	0.740	(0.021)	0.190	(0.021)	0.070	(0.003)
	4	0.722	(0.007)	0.188	(0.007)	0.090	(0.003)
	3	0.642	(0.007)	0.250	(0.007)	0.108	(0.003)
	2	0.595	(0.007)	0.279	(0.007)	0.126	(0.003)
	1	0.615	(0.006)	0.251	(0.006)	0.134	(0.005)
Parameter	ADL(j)	Estimate Extreme Profile-k (sd)					
		$k = 1$		$k = 2$		$k = 3$	
β_{0**}	1 (EAT)	-9.038	(0.364)	-3.085	(0.061)	-0.118	(0.043)
	2 (BED)	-7.115	(0.162)	-1.710	(0.055)	3.533	(0.148)
	3 (MOB)	-5.365	(0.096)	-0.733	(0.045)	5.699	(0.267)
	4 (DRS)	-8.042	(0.228)	-2.230	(0.056)	1.999	(0.082)
	5 (BTH)	-4.457	(0.073)	-0.205	(0.039)	6.159	(0.287)
	6 (TLT)	-6.638	(0.152)	-1.735	(0.050)	2.457	(0.099)
β_{1**}	1 (EAT)	0.362	(0.019)	0.345	(0.009)	0.098	(0.006)
	2 (BED)	0.396	(0.010)	0.540	(0.013)	0.286	(0.013)
	3 (MOB)	0.347	(0.007)	0.510	(0.012)	0.419	(0.021)
	4 (DRS)	0.394	(0.013)	0.454	(0.011)	0.199	(0.008)
	5 (BTH)	0.292	(0.006)	0.414	(0.009)	0.438	(0.023)
	6 (TLT)	0.336	(0.010)	0.465	(0.011)	0.229	(0.010)
$Age_{1/2}$	1 (EAT)	104.964	(0.492)	88.956	(0.174)	81.215	(0.483)
	2 (BED)	97.991	(0.176)	83.168	(0.092)	67.651	(0.184)
	3 (MOB)	95.454	(0.147)	81.438	(0.086)	66.384	(0.155)
	4 (DRS)	100.435	(0.242)	84.914	(0.109)	69.943	(0.204)
	5 (BTH)	95.241	(0.158)	80.495	(0.094)	65.939	(0.166)
	6 (TLT)	99.784	(0.246)	83.731	(0.099)	69.287	(0.185)

Table 5.3: Posterior means for parameters of interest for basic model with $K = 3$ extreme profiles. Numbers between parenthesis are posterior standard deviations.

Parameter	Estimate
$\alpha_{0(5)}$	0.213 (0.016)
$\alpha_{0(4)}$	0.195 (0.013)
$\alpha_{0(3)}$	0.196 (0.010)
$\alpha_{0(2)}$	0.266 (0.010)
$\alpha_{0(1)}$	0.282 (0.010)

Parameter	Cohort	Estimate Extreme Profile-k (sd)							
		$k = 1$		$k = 2$		$k = 3$		$k = 4$	
ξ	5	0.592	(0.066)	0.247	(0.066)	0.107	(0.006)	0.054	(0.003)
	4	0.644	(0.012)	0.173	(0.011)	0.122	(0.005)	0.061	(0.003)
	3	0.559	(0.009)	0.224	(0.008)	0.134	(0.005)	0.084	(0.003)
	2	0.501	(0.008)	0.275	(0.008)	0.117	(0.004)	0.107	(0.003)
	1	0.518	(0.007)	0.280	(0.007)	0.102	(0.005)	0.100	(0.005)
Parameter	ADL(j)	Estimate Extreme Profile-k (sd)							
		$k = 1$		$k = 2$		$k = 3$		$k = 4$	
β_{0**}	1 (EAT)	-10.817	(0.648)	-5.094	(0.138)	-1.221	(0.069)	-0.373	(0.051)
	2 (BED)	-8.061	(0.235)	-4.621	(0.152)	0.914	(0.086)	4.482	(0.306)
	3 (MOB)	-6.090	(0.139)	-3.135	(0.109)	1.605	(0.088)	7.741	(0.575)
	4 (DRS)	-9.630	(0.429)	-5.072	(0.155)	0.199	(0.088)	1.850	(0.136)
	5 (BTH)	-5.027	(0.115)	-2.242	(0.084)	1.873	(0.084)	8.536	(0.787)
	6 (TLT)	-7.616	(0.247)	-4.271	(0.127)	0.659	(0.089)	2.276	(0.175)
β_{1**}	1 (EAT)	0.417	(0.031)	0.452	(0.016)	0.425	(0.015)	0.036	(0.006)
	2 (BED)	0.418	(0.014)	0.741	(0.025)	0.635	(0.023)	0.301	(0.025)
	3 (MOB)	0.365	(0.010)	0.676	(0.020)	0.541	(0.018)	0.508	(0.043)
	4 (DRS)	0.441	(0.022)	0.655	(0.021)	0.564	(0.020)	0.133	(0.014)
	5 (BTH)	0.302	(0.009)	0.563	(0.017)	0.451	(0.013)	0.548	(0.059)
	6 (TLT)	0.354	(0.014)	0.634	(0.020)	0.550	(0.019)	0.161	(0.017)
$Age_{1/2}$	1 (EAT)	105.988	(0.609)	91.264	(0.191)	82.875	(0.194)	90.966	(3.246)
	2 (BED)	99.298	(0.224)	86.236	(0.100)	78.560	(0.127)	65.107	(0.377)
	3 (MOB)	96.703	(0.181)	84.640	(0.098)	77.034	(0.137)	64.752	(0.280)
	4 (DRS)	101.840	(0.315)	87.741	(0.114)	79.649	(0.153)	66.030	(0.646)
	5 (BTH)	96.631	(0.202)	83.980	(0.101)	75.846	(0.143)	64.408	(0.335)
	6 (TLT)	101.497	(0.334)	86.740	(0.111)	78.802	(0.147)	65.820	(0.611)

Table 5.4: Posterior means for parameters of interest for basic model with $K = 4$ extreme profiles. Numbers between parenthesis are posterior standard deviations.

Parameter	Estimate
$\alpha_{0(5)}$	0.227 (0.018)
$\alpha_{0(4)}$	0.186 (0.010)
$\alpha_{0(3)}$	0.241 (0.009)
$\alpha_{0(2)}$	0.252 (0.009)
$\alpha_{0(1)}$	0.254 (0.010)

Parameter	Cohort	Estimate Extreme Profile-k (sd)				
		$k = 1$	$k = 2$	$k = 3$	$k = 4$	$k = 5$
ξ	5	0.139 (0.032)	0.647 (0.040)	0.110 (0.020)	0.070 (0.004)	0.036 (0.002)
	4	0.591 (0.034)	0.170 (0.031)	0.129 (0.006)	0.072 (0.004)	0.038 (0.002)
	3	0.528 (0.013)	0.205 (0.009)	0.132 (0.008)	0.072 (0.004)	0.063 (0.003)
	2	0.465 (0.014)	0.247 (0.009)	0.130 (0.009)	0.076 (0.004)	0.081 (0.003)
	1	0.468 (0.014)	0.259 (0.012)	0.110 (0.006)	0.072 (0.004)	0.091 (0.004)
Parameter	ADL(j)	Estimate Extreme Profile-k (sd)				
		$k = 1$	$k = 2$	$k = 3$	$k = 4$	$k = 5$
β_{0**}	1 (EAT)	-11.502 (2.284)	-5.852 (0.220)	-2.542 (0.126)	0.523 (0.104)	-1.602 (0.102)
	2 (BED)	-10.162 (0.437)	-6.995 (0.304)	-1.174 (0.181)	4.351 (0.247)	1.560 (0.131)
	3 (MOB)	-7.774 (0.245)	-4.741 (0.241)	-0.039 (0.146)	4.876 (0.259)	32.043 (6.965)
	4 (DRS)	-11.211 (0.749)	-7.126 (0.314)	-1.685 (0.163)	3.573 (0.227)	-0.058 (0.103)
	5 (BTH)	-6.300 (0.194)	-3.282 (0.169)	0.565 (0.123)	4.944 (0.224)	26.035 (6.995)
	6 (TLT)	-8.682 (0.414)	-5.964 (0.241)	-1.122 (0.166)	4.029 (0.230)	0.363 (0.093)
β_{1**}	1 (EAT)	0.441 (0.216)	0.481 (0.018)	0.508 (0.020)	0.364 (0.018)	-0.110 (0.010)
	2 (BED)	0.589 (0.026)	0.962 (0.047)	1.059 (0.057)	0.698 (0.030)	-0.160 (0.019)
	3 (MOB)	0.506 (0.022)	0.791 (0.033)	0.819 (0.037)	0.633 (0.028)	-2.122 (0.480)
	4 (DRS)	0.585 (0.032)	0.844 (0.035)	0.867 (0.042)	0.671 (0.029)	-0.132 (0.013)
	5 (BTH)	0.412 (0.017)	0.616 (0.022)	0.595 (0.023)	0.577 (0.022)	-1.754 (0.486)
	6 (TLT)	0.459 (0.019)	0.782 (0.035)	0.858 (0.039)	0.685 (0.029)	-0.118 (0.013)
$Age_{1/2}$	1 (EAT)	98.636 (382.190)	92.184 (0.402)	85.004 (0.261)	78.570 (0.242)	65.364 (0.796)
	2 (BED)	97.247 (0.522)	87.277 (0.314)	81.114 (0.201)	73.766 (0.185)	90.001 (2.697)
	3 (MOB)	95.367 (0.481)	85.997 (0.278)	80.050 (0.187)	72.303 (0.176)	95.177 (0.946)
	4 (DRS)	99.157 (0.532)	88.451 (0.327)	81.949 (0.227)	74.675 (0.191)	79.636 (1.018)
	5 (BTH)	95.291 (0.494)	85.329 (0.294)	79.051 (0.203)	71.437 (0.165)	94.916 (1.004)
	6 (TLT)	98.900 (0.611)	87.638 (0.328)	81.310 (0.211)	74.117 (0.175)	83.225 (1.595)

Table 5.5: Posterior means for parameters of interest for basic model with $K = 5$ extreme profiles. Numbers between parenthesis are posterior standard deviations.

			Extreme Profiles			
Age Range	Pattern	Observed	$K = 2$	$K = 3$	$K = 4$	$K = 5$
$65 - 70$	000000	21361	20004	20672	21070	21255
	111111	295	0	12	45	177
	010000	292	526	289	238	290
	111110	192	6	45	109	258
	000100	167	422	211	151	174
	010110	117	65	103	94	67
	010100	103	181	140	90	72
	000110	89	101	80	58	63
$70 - 79$	000000	34508	32177	33344	33550	33776
	010000	1143	1209	1388	1223	1227
	111111	1087	153	396	361	424
	111110	692	406	734	723	782
	000100	584	767	732	662	619
	010100	412	412	334	418	407
	010110	375	348	253	289	285
	110110	335	359	357	372	386
$79 - 80$	000000	15536	13423	13960	14371	14624
	111111	1734	801	811	989	1160
	010000	1178	1827	1339	1379	1237
	111110	1018	938	1002	1068	901
	000100	562	1219	869	832	712
	010100	536	410	533	506	553
	010110	529	265	373	354	437
	110110	461	362	442	449	468
$90 <$	000000	1311	397	745	909	906
	111111	1025	562	792	850	856
	111110	529	309	358	457	462
	010110	196	157	126	116	132
	010000	193	295	338	315	286
	010100	189	245	210	178	191
	110110	158	139	113	114	129
	000100	138	261	280	260	232
		X^2	123563	14938	7638	4646

Table 5.6: Observed and mean predictive posterior aggregated counts for the 10 most populated response patterns by age range, for the Grouped model with $K = 2, 3, 4, 5$. Within each age group the response patterns are sorted decreasingly according to observed frequencies. For each model the X^2 statistic has been computed from the complete $2^6 \times 4$ contingency table.

Chapter 6

Modeling Mortality

In this chapter I present an extension of the Basic Grade of Membership Trajectory Model that also includes the survival outcome as a response. This extension is specifically directed at the particularities of the NLTCS but can also be generalized to other settings.

There are many reasons to undertake this task. First, in the context of studies about disability, the study of survival times is of intrinsic interest. From a public policy point of view, how long an individual is going to survive has a direct impact on the allocation of resources, even more if different patterns of disability are correlated with different patterns of mortality.

Second, disability is known to be correlated with mortality in elderly years. Whichever the working mechanism involved is, it has been shown repeatedly that progression into disability goes together with an increased probability of death (e.g. Ferrucci et al. 1996; Connor 2006; White 2008). As we discussed in Section 2.1.1, because of this correlation, ignoring the missing data due to death without further consideration is a clear violation to the Missing At Random assumption introduced in Section 2.1.1 (Rubin, 1976). This can be very problematic since death is the most important cause of attrition from the NLTCS sample (White, 2008).

Third, if we accept that patterns of disability and of mortality go together, information about survival times can help to make up for scarcity of data points needed to reconstruct certain regions of some disability patterns. For instance, consider the case of a young non-disabled individual that appeared in just one wave of the NLTCS because of their death. The extra information about this early death can in principle help to differentiate this subject from other healthy individuals, with exactly the same disability outcomes, but with longer survival times. Models that do not consider joint-estimation of survival and disability, like the those presented in Section 2 or in Connor (2006) would assume all these individuals exchangeable.

Kurland and Heagerty (2005) and Kurland et al. (2009) introduce a general framework to approach the problem of longitudinal modeling with dropout caused by death that can help to clarify this situation. Consider the longitudinal response of interest, Y_i, and the survival times, S_i. Since the length of Y_i is determined by S_i, modeling of Y_i will be influenced by the distribution of S_i, either implicitly or explicitly. In terms of modeling we can either directly model the joint distribution of (Y_i, S_i), $p(X_i, S_i)$, or to propose models for the components of the different possible factorizations of this distribution: $p(Y_i, S_i) = p(Y_i|S_i)p(S_i) = p(S_i|Y_i)p(Y_i)$.

Thus far, the proposed models for the disability responses the “Basic” (Chapter 2) and “Grouped” (Chapter 4 models, have been specified for the unconditional version, $p(Y_i)$, which correspond to the idea of an "immortal cohort" (or several, according to models in Chapter 4). This approach would only be adequate by itself in situations in which people do not die during he period of study, or when the values of the response have well defined values after death. As none of these conditions are met in the NLTCS, the estimates will reflect the averaging of $p(Y_i|S_i)$ over the underlying distribution of S_i (Kurland et al., 2009).

The approach that I take in this chapter will be to consider a full joint model for disability and mortality, or to model $p(Y_i, S_i)$, in Kurland et al. (2009) terminology). To this end, I rely on the Grade of Membership ideas, postulating *conditional* independence between disability and survival times, given the membership vector of each individual. In other words, I use the membership vector to decouple the dependence between disability and survival.

6.1 Modeling Mortality on the NLTCS

Analyses of the NLTCS that combine longitudinal disability patterns with mortality information have been scarce. Besides the inherent technical difficulties, this is likely due to the lack of reliable information about death in the original NLTCS sample. As commented in Chapter 1, researchers at Duke university seek to fill this void by linking the NLTCS registers with the Medicare records, creating a file (CMS) with the dates of birth of all subjects and, for those deceased, their date of death as of February, 2006 (Stallard, 2005; Connor, 2006). This combined data file has been analyzed in Stallard (2005), Connor (2006), White (2008) and Manton et al. (2008).

Connor (2006) performed an analysis of patterns of mortality in the NLTCS based on estimates of disability obtained using the Multivariate Latent Trajectory framework (see 1.3.2). To this end, he stratified the NLTCS data based on the exclusive-cluster membership estimates obtained from the MVLT framework, using disability data alone. Then, *conditional on that stratification*, he obtained within-cluster survival estimates using Kaplan-Meier curves.

Connor's analysis presented an interesting first approximation to the heterogeneous nature of survival distributions and their interplay with the disability outcomes, that has some similarities with the approach I present in the here. His methodology does, however, have some technical shortcomings. First, although it implicitly assumes an interplay between disability and mortality, the survival estimation is conditional on the clustering induced by the disability outcomes alone. This has the effect, in the best case, of wasting information that could be useful for improving the classification and, in the worst case, of biasing the estimates by not differentiating missing data due to death from other causes. Second, survival estimates were obtained without delayed entry correction (see Section 6.2 for details), and rather relied on discarding data, producing loss of inferential power and risk of selection bias.

Stallard (2005) used the sequence of GoM scores over time, obtained from his Longitudinal GoM transition model (where GoM scores were considered as descriptions of states of disability; see Section 1.3 in Chapter 1 for details), as covariates to plug-in into a hazards-based model for mortality. In this way he was able to investigate the relationship between different states of disability and the chances of survival. Similar to Connor (2006), this estimation was performed conditional on the estimates of disability, and therefore mortality information did not play any role in the estimates of disability. This approach was also taken in Manton et al. (2008), although no details about estimation were provided there.

White (2008) (see 1.3) used survival information in his analyses by explicitly modeling death as a special –absorbing– state, among those used to represent. In this way he was able to estimate probability transitions from specific disability states into death. The estimations was done simultaneously with the rest of the parameters of the model, including those that represented disability, effectively sharing information between those two phenomena.

6.2 Practical Aspects of Modeling Mortality for NLTCS subjects

6.2.1 Initial Considerations

By design, all subjects in the NLTCS are older than 65 years. Thus, any attempt to estimate the distribution of ages at the time of death needs to take into consideration that the observed survival times are conditional on the event of having survived past the age 65. This fact should not be problematic, since we are precisely interested in the survival distribution of elderly people, this is, the conditional distribution of survival times given that the subjects have already lived more than 65 years. From now on, any reference to the distribution of survival times will refer to this conditional version. For instance, any reference to the cdf of the survival distribution will be understood as

$$F(d) = \Pr(S \leq d | S \geq 65).$$

6.2.2 Dealing with Right Censoring

The fist difficulty that we face when trying to estimate F is that survival observations are right-censored, meaning that not everybody will have died by the time of the last assessment. Therefore our best knowledge of the survival outcome for those people is that it is greater than their age at the last assessment. This is a very well known problem and most survival analysis techniques are developed specifically to deal with it, e.g., Kalbfleisch and Prentice (2002).

Let d^* denote the uncensored survival time or age of a person at the time of their death and b the last known age if the person was still alive then. Let ν be the censoring indicators (i.e. $\nu = 1$ if the subject is dead and $\nu = 0$ if the subject is still alive) and d the censored version of the age at the time of dead, i.e. $d = d^* \cdot \nu + b(1 - \nu)$. Then,

$$\Pr[S \leq d|\nu, b, S \geq 65] = F(d)^{1-\nu} \left[1 - F(b)\right]^{\nu}.$$

If we assume some parametric form for the survival distribution, say $F_\theta = F$ with density $f_\theta(\cdot)$ with respect to some convenient measure on the real line, we can easily write down the likelihood,

$$L(\theta|d, \nu, b) = \prod_{i=1}^{N} f_\theta(d_i)^{1-\nu_i}[1 - F_\theta(b_i)]^{\nu_i},$$

and estimate the parameters.

In principle this likelihood depends on both the density and the c.d.f. of a distribution, but the estimation of the parameters θ is in many cases a straightforward exercise using a missing data approach. Considering separately the complete data likelihood,

$$p(d^*|\theta) = \prod_{i=1}^{N} f_\theta(d_i^*)$$

and the censoring mechanism

$$p(\nu_i|d_i^*, b_i) = \delta_0(\nu_i)^{I(b_i<d_i^*)} \times \delta_1(\nu_i)^{I(b_i\geq d_i^*)}$$

we can apply an algorithm that takes advantage of this structure, such as EM. or data augmentation strategies such as those in Tanner (1996).

6.2.3 Dealing with left truncation

One particularity of the NLTCS dataset is that, although by design we know that all subjects are older than 65 years, many of them entered the survey at very different ages. This feature is specially relevant in the case of the oversampling of "oldest old", where people were specifically selected because of their advanced age (see Section 1.1).

If we take the naïve approach of estimating the distribution F_θ using regular survival analysis approaches, designed just to handle right-censoring (e.g. regular Kaplan-Meier estimators), the resulting estimates will be biased. This is due to the fact that individuals that have entered the sample at an older age, must have already had to survive to that age in order to be in the sample, making the data look as if surviving to late ages were more of a common occurrence than it really is. This will produce an overestimation of the survival probabilities at late ages.

This is a well-known situation, sometimes referred to as "delayed entry" (Hougaard, 2000;

Kalbfleisch and Prentice, 2002), and many analysis techniques have been adapted to deal with it (see e.g. Kalbfleisch and Prentice 2002). The basis for these solutions lies in specifying the distribution of survival times conditional on the age of entry to the sample. In this way we are specifying the same underlying distribution for every individual (and therefore estimating the same set of parameters), but considering the fact that the survival times that we *observe* are conditional on being greater than the age at which the individual first entered the sample. Ignoring for the moment the censoring problem, let S be a random variable specifying the survival time in excess of 65 years, a the age at the time of entering the sample (this implicitly makes $a \geq 65$) and f_θ be the density of F_θ with respect to some measure on the real line. Then we have that

$$\Pr(S \leq d | S > a) = \int_a^d \frac{f_\theta(t)}{1 - F_\theta(a)} dt \times I(d > a).$$

Now we can construct the correct likelihood for this sample,

$$L(\theta | d, a) \propto \prod_{i=1}^{N} \frac{f_\theta(d_i)}{1 - F_\theta(a_i)} \times I(d_i > a_i), \tag{6.1}$$

from where, at least in principle, we can estimate the underlying parameters, θ.

6.3 A Weibull model With Truncation and Delayed Entry

From a parametric estimation standpoint, the likelihood in Equation (6.1) can be computationally inconvenient, given that the parameters, θ, appear in an expression that involves both a density, $f_\theta(\cdot)$, and its integral, $F_\theta(x) = \int_{65}^{x} f_\theta(x)dx$. Expressions like these seldom possess a closed form. Although this is a commonly encountered situation, and there are well known and more or less standard approaches like the ones described in Kalbfleisch and Prentice (2002), the final objective of this modeling exercise is to integrate it with models describing disability status, into a comprehensive joint model. Thus a simple parametric specification is desirable.

A practical alternative is the use of the Weibull distribution (Weibull, 1951). The c.d.f. of the Weibull distribution is given by the expression:

$$F_w(x) = \Pr[X \leq x] = 1 - \exp\left[-(\beta x)^{\alpha}\right], \tag{6.2}$$

and its density by

$$f_w(x) = \alpha\beta^{\alpha}x^{\alpha-1}\exp[-(x\beta)^{\alpha}], \tag{6.3}$$

for $x \geq 0$. β and α are positive inverse scale and shape parameters respectively.

The Weibull distribution is a versatile unimodal distribution that can be used to model general nonnegative random variables, similar to the well known Gamma distribution. Its use is very extended in survival analysis in part due to its simplicity, the existence of multiplicative closed forms for its density, survival function and hazards function, and the fact that its hazards function is monotonic (Kalbfleisch and Prentice, 2002; Ibrahim et al., 2001). This last property makes it a good candidate to model mortality in the NLTCS, since hazards in elderly population are usually increasing (Stallard, 2005).

To illustrate, Figure 6.1 shows the density of a Weibull distribution for different values of the shape parameter. Values of $\alpha < 1$ specify decreasing hazards, while values $\alpha > 1$, increasing hazards (Kalbfleisch and Prentice, 2002).

If we substitute $f_\theta(x)$ and F_θ in Equation (6.1) by the Weibull density (Equation 6.3) and

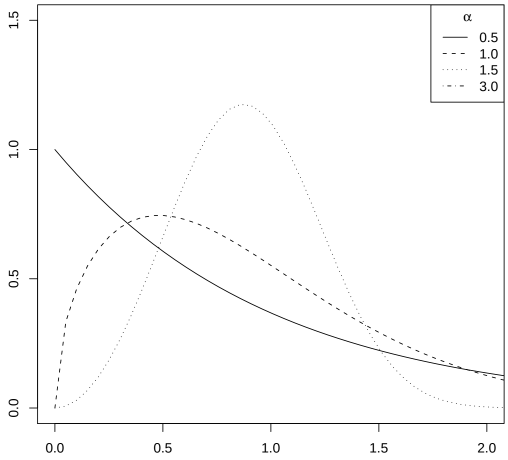


Figure 6.1: Density of the Weibull distribution for different values of the shape parameter α ($\beta = 1$).

c.d.f. (Equation 6.2), this likelihood reduces to the simple multiplicative expression:

$$L(\beta, \alpha | d, a) = \prod_{i=1}^{N} \alpha \beta^{\alpha} d_i^{\alpha - 1} \exp\left[\beta^{\alpha}(a_i^{\alpha} - d_i^{\alpha})\right] \times I(a_i < b_i). \tag{6.4}$$

From here, we can complete a Bayesian specification by choosing priors for α and β, with support on the interval $(0, \infty)$. Given that both α and β must be nonnegative, a reasonably flexible specification can be obtained by making

$$\begin{aligned} \alpha &\sim \text{Gamma}(a_{\alpha}, b_{\alpha}), \\ \beta &\sim \text{Gamma}(a_{\beta}, b_{\beta}). \end{aligned} \tag{6.5}$$

6.3.1 Estimation Using MCMC

The joint posterior of the parameters in the model specified through the likelihood in Equation (6.4) and priors in Equation (6.5), without considering the censoring is

$$p(\beta, \alpha | d^*, a) \quad \propto \quad p(\alpha)p(\beta) \prod_{i=1}^{N} \alpha\beta^\alpha d_i^{*\alpha-1} \exp\left[\beta^\alpha (a_i^\alpha - d_i^{*\alpha})\right] \times I(a_i < d_i^*) \tag{6.6}$$

with $p(\alpha) = \text{Gamma}(\alpha | a_\alpha, b_\alpha)$, $p(\beta) = \text{Gamma}(\beta | a_\beta, b_\beta)$.

To consider right censoring, we can apply a combination of a data-augmentation algorithm (Tanner, 1996) and a Metropolis-Hastings algorithm to sample from the posterior of (α, β). Such an algorithm can be implemented as follows

1. **Sampling $d_i^*|...$ (Imputation Step)** For all $i = 1, ..., N$, sample d_i^* according to

$$d_i^* = \begin{cases} d_i & \text{if } \nu_i = 1, \\ \text{Draw} \quad d_i^* \sim \text{LTWeib}(\alpha, \beta, a_i) & \text{if } \nu_i = 0, \end{cases}$$

   where $\text{LTWeib}(\alpha, \beta, a)$ is the left-truncated (at age a) Weibull distribution, with density

$$p(x|\alpha, \beta, a) = \alpha\beta^\alpha x^{\alpha-1} \exp\left[\beta^\alpha (a^\alpha - x^\alpha)\right] I(x > a).$$

   To obtain a sample, d^*, from this distribution we can use a variant of the inverse c.d.f. method (Devroye, 1986), taking advantage of the existence of closed forms for the c.d.f. and inverse c.d.f. functions:

   (a) sample $u \sim \text{Unif}\,(F(a|\alpha, \beta), 1)$

   (b) make $d = F^{-1}(u|\alpha, \beta) = \frac{1}{\beta}\left[\log(1-u)\right]^{1/\alpha}$, where $F(\cdot)$ is the cdf of the Weibull distribution and $F^{-1}(\cdot)$ is its inverse.

2. **Sampling from** $(\alpha, \beta)|...$: From Equation (6.6), the expression for the full conditional distribution of (α, β) is

$$\begin{aligned}
&p(\alpha, \beta|...) \\
&\quad \propto p(\alpha, \beta) \times \prod_{i=1}^{N} \mathrm{LTWeib}(d_i^*|\alpha, \beta, a_i) \\
&\quad \propto p(\alpha)p(\beta) \times (\alpha\beta^{a})^N \exp\left[\alpha \sum_{i=1}^{N} \log d_i^* + \beta^{\alpha} \sum_{i=1}^{N} ({a_i}^{\alpha} - d_i^{*\alpha})\right] \\
&\quad \propto \alpha^{a_\alpha - 1} \exp\left(-b_\alpha \alpha\right) \beta^{a_\beta - 1} \exp\left(-b_\beta \beta\right) \times (\alpha\beta^{a})^N \exp\left[\alpha \sum_{i=1}^{N} \log d_i^* + \beta^{\alpha} \sum_{i=1}^{N} ({a_i}^{\alpha} - d_i^{*\alpha})\right] \\
&\quad = \alpha^{N + a_\alpha - 1} \beta^{N\alpha + a_\beta - 1} \exp\left[\alpha\left(\sum_{i=1}^{N} \log d_i^* - b_\alpha\right) + \beta^{\alpha} \sum_{i=1}^{N} ({a_i}^{\alpha} - d_i^{*\alpha}) - b_\beta \beta\right].
\end{aligned}$$

Since this expression does not have any recognizable form, we can use the following Metropolis-Hastings step:

(a) (proposal step) Sample a proposal value (α^*, β^*) from

$$\begin{aligned}
\log \alpha^* &\sim N\left(\log \alpha, \sigma_\alpha^{*2}\right), \\
\log \beta^* &\sim N\left(\log \beta, \sigma_\beta^{*2}\right).
\end{aligned}$$

(b) (Acceptance step) The proposal distribution is not symmetric, so we will need to compute both the Metropolis ratio and the Hastings correction for asymmetry.

Compute

$$\begin{aligned}
r_M =& \frac{p(\alpha^*, \beta^* | ....)}{p(\alpha, \beta | ....)} \\
=& \left(\frac{\alpha^*}{\alpha}\right)^{N+a_\alpha-1} \frac{\beta^{*\alpha^* N + a_\beta - 1}}{\beta^{\alpha N + a_\beta - 1}} \\
& \times \exp\Bigg[(\alpha^* - \alpha)\left(\sum_{i=1}^{N} \log d_i - b_\alpha\right) + \beta^{*\alpha^*}\left(\sum_{i=1}^{N}(a_i^{\alpha^*} - d_i^{\alpha^*})\right) \\
& \qquad - \beta^\alpha \left(\sum_{i=1}^{N}(a_i^\alpha - d_i^\alpha)\right) - b_\beta\left(\beta^* - \beta\right)\Bigg], \\
r_H =& \frac{q(\alpha, \beta | \alpha^*, \beta^*)}{q(\alpha^*, \beta^* | \alpha, \beta)} = \frac{\text{Lognormal}(\alpha | \log \alpha^*, \sigma_\alpha^{*2})}{\text{Lognormal}(\alpha^* | \log \alpha, \sigma_\alpha^{*2})} \cdot \frac{\text{Lognormal}(\beta | \log \beta^*, \sigma_\beta^{*2})}{\text{Lognormal}(\beta^* | \log \beta, \sigma_\beta^{*2})} \\
=& \frac{\alpha^*}{\alpha}\frac{\beta^*}{\beta}.
\end{aligned}$$

Then make

$$\begin{aligned}
r =& r_M \times r_H \\
=& \left(\frac{\alpha^*}{\alpha}\right)^{N+a_\alpha} \frac{\beta^{*\alpha^* N + a_\beta}}{\beta^{\alpha N + a_\beta}} \\
& \times \exp\Bigg[(\alpha^* - \alpha)\left(\sum_{i=1}^{N} \log d_i - b_\alpha\right) + \beta^{*\alpha^*}\left(\sum_{i=1}^{N}(a_i^{\alpha^*} - d_i^{\alpha^*})\right) \\
& \qquad - \beta^\alpha \left(\sum_{i=1}^{N}(a_i^\alpha - d_i^\alpha)\right) - b_\beta\left(\beta^* - \beta\right)\Bigg].
\end{aligned}$$

Finally, update the current value of the pair (α, β), from step m to $m+1$ according to

$$(\alpha, \beta)^{(m+1)} = \begin{cases} (\alpha^*, \beta^*) & \text{with probability } \min\{r, 1\}, \\ (\alpha, \beta)^{(m)} & \text{with probability } 1 - \min\{r, 1\}. \end{cases}$$

6.3.2 Testing the Weibull survival model with NLTCS mortality data

To test the algorithm in the previous section, I applied it to a subset of the NLTCS consisting in 6000 randomly selected individuals. I certainly do not expect this simplistic model to provide a realistic fit of these data. This computation exercise is just for testing the algorithm and, in particular to evaluate how well it handles the censoring and delayed entry. I introduce a more realistic model in the next section.

I have selected priors for α and β as

$$\alpha \sim \text{Gamma}(1,1) \quad \text{and} \quad \beta \sim \text{Gamma}(1, 0.1),$$

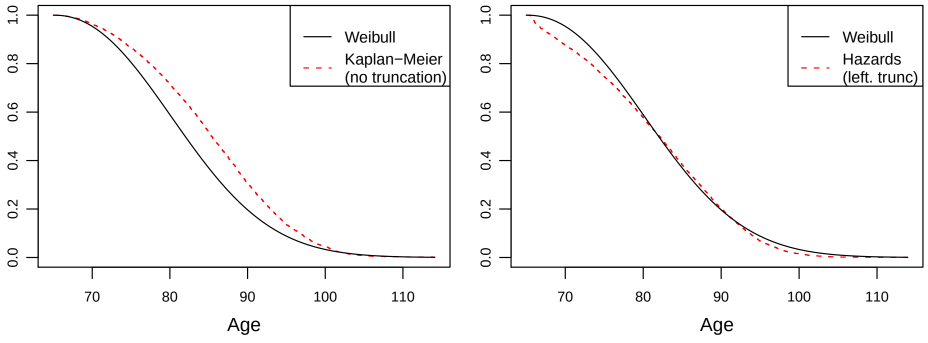
which can be considered diffuse, but with a realistic scale and shape to model human survival times in excess of 65 years.

For the tuning parameters, σ_α and σ_β, I have selected the values $\sigma_\alpha = \sigma_\beta = 0.01$, which produce an acceptance rate of approximately 45% in the Metropolis-Hastings step. With these values, the sampler converges quickly, after just 500 iterations, so I ran $10,000$ long chains and discarded the first 1000. No subsampling was necessary.

The estimated posterior mean and standard deviation of the Weibull parameters are

$$\hat{\alpha} \text{ (sd)} = 2.20 \text{ } (0.032),$$
$$\hat{\beta} \text{ (sd)} = 0.0498 \text{ } (0.000388).$$

Figure 6.1 shows the estimated posterior survival curve overlaid for comparison with the Kaplan-Meier survival estimate without left truncation correction, and with a hazards-



(a) (b)

Figure 6.2: Estimated Weibull survival curve for NLTCS subjects compared with no deleyed-entry corrected Kaplan-Meier estimate (left panel) and with hazards based estimate with delayed entry correction (right panel).

based estimate (Cox model without covariates, in the left panel) with delayed entry correction. We can visually assess the effect of the left truncation correction. If we take the corrected hazards-based estimate as a benchmark, we see that not correcting for delayed entry (Kaplan-Meier curve) has the expected effect of overestimating the probability of survival. Comparing the hazards-based corrected curve with the Weibull model, we see that they are very close to each other, meaning that the correction in the Weibull model is handling the delayed entry bias at least in a comparable way. The main discrepancy in shape between the two curves is near the origin, where the Weibull model does not show as sharp decay as the nonparametric estimate does. This is likely due to model misspecification. We have to remember the strong parametric form of the Weibull distribution, which is a smooth curve that is uniquely determined by just two scalar parameters. A more sophisticated approach will be necessary to correct this bias.

6.4 Constructing a Joint Model for Mortality and Disability

The objective of this section is to construct a comprehensive probability model that can handle both survival and disability outcomes, considering them as distinct aspects of the same underlying processes.

To this end, I use the partial membership structure as a way of connecting both phenomena. I therefore specify ideal profiles that describe the progression of survival probabilities as well as survival probabilities. By assuming conditional independence between those distributions given the membership vector, I make the individual membership serve as a way of explaining the dependence structure between the two.

I begin by proposing an extension of the simple Weibull model from the previous section, aimed to relax the strong parametric constrains imposed by the choice of the Weibull distribution. This involves introducing an individual-level mixture structure, using GoM scores. Then I combine it with the Basic GoM trajectory model from Chapter 2 by letting the (sub)models for disability outcomes and the survival times to share the same membership vector as a way of linking the two pieces.

6.4.1 A GoM version of the simple Weibull survival model

Assuming the existence of K extreme profiles, we model each of these extreme profiles with the simple Weibull model:

$$p(d|g_k = 1) = \text{Weibull}(d|\alpha_k, \beta_k) \tag{6.7}$$

and then model each individual as a weighted (by the membership vector) combination of the unique profiles:

$$p(d_i|g_i) = \sum_{k=1}^{K} g_{ik} \text{Weibull}(d_i|\alpha_k, \beta_k).$$

Again assuming independence conditional on the membership vector, we construct the joint distribution:

$$p(d|g) = \prod_{i=1}^{N} \sum_{k=1}^{K} g_{ik} \text{Weibull}(d_i|\alpha_k, \beta_k),$$

which can be rewritten using a latent class representation

$$p(d|g) = \prod_{i=1}^{N} \sum_{z \in Z} \prod_{k=1}^{K} \left(g_{ik} \text{Weibull}(d_i|\alpha_k, \beta_k) \right)^{I(z=k)},$$

where $Z = \{1, 2, ..., K\}$. It is easy to see that this joint distribution is also the distribution of variates generated by the following process:

1. for each $i = 1, ..., N$
    (a) draw $z_i \sim \text{Discrete}[g_{i1}, g_{i2}, ..., g_{iK}]$,
    (b) draw $d_i \sim \text{Weibull}(\alpha_{z_i}, \beta_{z_i})$.

A quick derivation will show us that when we consider the membership vectors to be samples from a common distribution, G, this specification becomes nothing else than a simple (constrained) mixture of Weibulls model:

$$\begin{aligned}
p(d|\theta) &= \int_{\Delta} \sum_{k=1}^{K} f_k(d) G(dg) \\
&= \int_{\Delta} \sum_{k=1}^{K} g_k \text{Weibull}(d|\alpha_k, \beta_k) G(dg) \\
&= \sum_{k=1}^{K} \pi_k \text{Weibull}(d|\alpha_k, \beta_k), \qquad (6.8)
\end{aligned}$$

where $\pi_k = E_G[g_k]$ is the expected value of the k-th component of g ($\pi_k = \xi_k$, for the Dirichlet distribution).

Although this specification may look trivial by itself (and it is nothing more than a hierarchical mixture model), its power will become evident when we combine it with whole trajectory model, considering shared membership vectors for both the disability trajectories and mortality, and make each component part of an extreme profile that also characterizes disability trajectories.

Adding delayed-entry correction (Equation 6.1) to this formulation, we obtain the expression

$$\begin{aligned}
p(d_i|\theta, g_i, a_i) &= \frac{f_\theta(d)}{1 - \int_0^{a_i} f_\theta(x) dx} \\
&= \frac{\sum_{k=1}^{K} g_{ik} \text{Weibull}(d_i|\alpha_k, \beta_k)}{1 - \int_0^{a_i} \sum_{k=1}^{K} g_{ik} \text{Weibull}(x|\alpha_k, \beta_k) dx} \\
&= \frac{\sum_{k=1}^{K} g_{ik} \text{Weibull}(d_i|\alpha_k, \beta_k)}{\sum_{k=1}^{K} g_{ik} \left(1 - \int_0^{a_i} \text{Weibull}(x|\alpha_k, \beta_k) dx\right)}. \qquad (6.9)
\end{aligned}$$

This expression will not admit a latent class representation like the one in Equation (6.8), due to the sum in the denominator. As a consequence, the estimation strategies based on data augmentation that we have applied previously are no longer feasible. This added complication is an unfortunate side effect of having to perform the truncation in the mixed distribution (i.e., after generating the variable) and not in the extreme profiles.

We obtain a compromise solution by replacing the distribution in Equation (6.9) with

$$\begin{aligned} p(d_i|g_i, a_i) &= \sum_{k=1}^{K} g_{ik} \frac{\text{Weibull}(d_i|\alpha_k, \beta_k)}{1 - \int_0^{a_i} \text{Weibull}(x|\alpha_k, \beta_k) dx} \\ &= \sum_{k=1}^{K} g_{ik} \text{LTWeib}(d_i|\alpha_k, \beta_k, a_i), \end{aligned} \tag{6.10}$$

by truncating each extreme profile and *then* mixing the distributions according to the membership scores. This expression has the advantage of being a true individual-level mixture, and therefore to define a probability measure. In addition, it admits a latent class representation. However, how well this surrogate expression approximates the true intended distribution of the model is something that we will have to evaluate. I come back to this question from an empirical point of view at the end of the next Chapter, when we apply the model to the actual NLTCS data.

6.4.2 Combining the Basic GoM Trajectory Model and Weibull Mortality Model

My main objectives when modeling the survival distribution using the NLTCS data are: 1) to understand the mortality patterns and their relationship to the disability patterns and 2) to supplement or complement the disability data with the survival data, in order to achieve a better classification of the individuals. These two objectives have the implicit requirement that, in some sense, the extreme profiles encode not only the typical disability curves, but also typical mortality distributions. If we proceed in this way, the prototypical individuals described by the extreme profiles will be described by their ways of aging and their probability of surviving by more than a given number of years.

As I outlined at the beginning of this section, for the construction of the joint model I assume:

1. Each extreme profile specifies both the trajectories of disability *and* the mortality distribution of the ideal individuals.
2. For each individual i, there is only one K-dimensional membership vector, g_i, that will apply to both the survival and the disability sub-models.
3. Given the membership vector, the survival distribution of an individual will be considered independent of their disability distribution.

The last assumption on the list is perhaps the most important. It states that the relationship between mortality and disability could be *fully* explained by the extreme profile partial membership.

Making these assumptions, we can write the joint distribution of mortality and disability conditional on the membership vector. Let $h_k(d)$ be the density of the survival distribution for the extreme profile k. Then

$$\begin{aligned} p(y_i, d_i|g_i) &= p(y_i|g_i) \times \sum_{k=1}^{K} g_{ik} h_k(d_i) \\ &= \underbrace{\left(\prod_{j=1}^{J} \prod_{t=1}^{T} \sum_{k=1}^{K} g_{ik} f_k(y_{ijt}|X_{it}) \right)}_{\text{Original Basic Model}} \times \sum_{k=1}^{K} g_{ik} h_k(d_i) \\ &= \sum_{z^* \in Z^*} \left[g_{iz'} h_{z'}(d_i) \prod_{j=1}^{J} \prod_{t=1}^{T} g_{iz_{jt}} f_{z_{jt}}(y_{ijt}|X_{it}) \right]. \end{aligned} \tag{6.11}$$

Where $Z^* = Z' \times Z$ with

$$\begin{aligned} Z' &= \{1, 2, ..., K\} \\ Z &= \{1, 2, ..., K\}^{J \times K} \end{aligned}$$

$z^* = (z, z')$, $z' \in Z'$ and $z \in Z$. Comparing expression in Equation (6.11) with the Basic model from Chapter 2 (Equation 2.5) we see that vectors $z^* \in Z^*$ are just an augmentation of vector in Z in Equation (2.5), with a component to accommodate the survival sub-model.

Using Equation (6.11), we can again write the likelihood of the joint model using the augmented data representation introduced in Chapter 2, Section 2.6:

$$p(y, d, z^*|g) = \prod_{k=1}^{K} (g_{ik} h_k(d_i))^{I(z'_i = k)} \times \prod_{j=1}^{J} \prod_{t=1}^{T} \prod_{k=1}^{K} [g_{ik} f_k(y_{ijt}|X_{it})]^{I(z_{ijt} = k)}. \tag{6.12}$$

This representation, is the basis for the construction of the estimation algorithm. Note that the representation in Equation (6.12) also falls into the general class of Mixed Membership models as described in Erosheva (2002).

6.4.3 Detailed Implementation

As I have proceeded in the previous sections, I implement this expanded model by selecting suitable disability trajectory functions, a survival distribution and priors.

For the Basic Model section I use the exact same specification proposed for the Basic model, this is

$$f_k(y|X_{it}) = \text{Bernoulli}\left[y|p = \text{logit}^{-1}\left(\beta_{0j|k} + \beta_{1j|k} \times Age_{it}\right)\right],$$

with the priors

$$\begin{aligned} p(g_{i*}|\alpha) &= \text{Dirichlet}(g_{i*}|\alpha_1, \alpha_2, ..., \alpha_k), \\ p(\alpha_0) &= \text{Gamma}(\alpha_0|\tau, \eta), \\ p(\xi) &= \text{Dirichlet}(\xi|\mathbf{1}_K) \quad \text{(Uniform on } \Delta_{K-1}). \end{aligned}$$

For the survival density function, $h(\cdot)$ I use the Weibull specification from Equation (6.10),

$$h_k(d_i|a_i) = \text{LTWeib}(d_i|\alpha_{dk}, \beta_{dk}, a_i),$$

where a_i is the left truncation point, as we discussed in Section 6.2.3. Similar to the construction of the simple Weibull model in Section 6.3, we will use the priors

$$\alpha_{dk} \overset{iid}{\sim} \text{Gamma}(a_\alpha, b_\alpha) \quad \text{and} \quad \beta_{dk} \overset{iid}{\sim} \text{Gamma}(a_\beta, b_\beta).$$

This specification does not include the handling of right censoring of survival times, which will be considered as part of the sampling strategy in the construction of the MCMC sampler in Section 6.4.5.

6.4.4 Missing Values

One of the objectives behind the construction of this Joint model under construction is to handle the non-ignorability of missing data when it is due to death.

The formulation in Equation (6.12) assumes that the full $T \times J$ dimensional vector of disability measurements is observed and available for estimation. This is certainly not only not the case, but a logical impossibility, as in several cases it would require to obtain disability measurements from deceased people.

To investigate the effect of ignoring the missing data under this formulation, I use a more abstract representation of the structure of the joint model, where the main assumption, the conditional independence of disability and survival processes, is isolated from other considerations. For now, I ignore the values of disability responses after death.

In the most basic terms, the joint model consists of three basic groups of parameters: the disability parameters, $\theta_d = (\beta_0, \beta_1)$; the survival parameters, $\theta_s = (\beta_d, \alpha_d)$ and; the mixed membership parameters, $\theta_c = (\alpha, g)$. The joint model is based on the idea of using the basic disability model likelihood, $f(y|\theta_d, \theta_c)$, and assuming conditional independence between these disability outcomes, y, and the survival outcomes, d, given the membership vector, $g \in \theta_c$. We label the model for the survival outcomes $h(d|\theta_s, \theta_c)$. Then, according to the joint specification,

$$\Pr(Y_i = y_i, S_i = d_i | X, \theta_d, \theta_s, \theta_c) = f(y_i | X_i, \theta_d, \theta_c) \cdot h(y_i | \theta_s, \theta_c).$$

Let us now assume that there are only three reasons why a person, i, would not have a measurement at time t. The person could be either

1. younger than 65 years ($65 > Age_{it}$) or
2. dead ($d_i < Age_{it}$) or
3. just happened to not have answered the survey for reasons that do not depend on either the disability response, y_i or the parameters of the model.

Note that these assumptions ignore the situation where the nonresponse depends on the unobserved disability outcome, like when the respondent is too disabled to be able to answer the survey. Although this might be a source of legitimate concern, this effect is likely to not being important given the design of the NLTCS, which specifically addresses this situation through the use of proxy respondents.

Let us define the missing data indicator for each individual as $m_{it} = I(t \in M_i)$. In the NLTCS this is an observed random variable. With this definition, the simplest way in which we can model the situation described in item 3 in the list of possible reasons for not being in the sample at time t (when $m_{it} = 1$) is by assuming that the non-response indicator comes from a common distribution $m_{it} \sim \phi$.

Now, let us define the set of indexes $\Xi_i = \Xi(X_i, S_i) = \big\{t : Age_{it} \in [65, S_i]\big\}$. This set contains the indexes of the waves in which the individual i could *in principle* be selected.

Then, the distribution of the missing data indicator, m_{it}, conditional on the age covariate and the survival time is

$$\begin{aligned} p(m_{it}|X_i, S_i) =& \delta_o(m_{it})^{I(t \notin \Xi)} \phi(m_{it})^{I(t \in \Xi)} \\ =& \delta_o(m_{it})^{I(t \notin \Xi)} \phi(1)^{I(t \in \Xi, m_{it}=1)} \phi(0)^{I(t \in \Xi, m_{it}=0)}. \end{aligned} \tag{6.13}$$

The factorization in the last line of Equation (6.13) just expresses the idea that sometimes, although an individual should in principle be in the sample, with probability $\phi(0)$ he or she just happens not to be included.

Dropping the individual index for simplicity of notation, the likelihood of the joint model considering the missing data indicator vector, m_i, is

$$\begin{aligned} p(y, m,& d|X, \theta_d, \theta_s, \theta_c) \\ &= h(d|\theta_s, \theta_c) \prod_{t=1}^{T} p(m_t|X, d) \prod_{j=1}^{J} f(y_{jt}|X_t, \theta_d, \theta_c) \\ &= h(d|\theta_s, \theta_c) \prod_{t \in \Xi} \left(\phi(m_t) \prod_{j=1}^{J} f(y_{jt}|X_t, \theta_d, \theta_c) \right) \times \prod_{t \notin \Xi} \left(\delta_0(m_t) \prod_{j=1}^{J} f(y_{jt}|X_t, \theta_d, \theta_c) \right) \\ &= h(d|\theta_s, \theta_c) \prod_{t \in M_i} \left(\phi(1) \prod_{j=1}^{J} f(y_{jt}|X_t, \theta_d, \theta_c) \right) \\ &\quad \times \prod_{t \notin M_i} \left(\phi(1)^{I(t \notin \Xi)} \delta_0(m_t)^{I(t \in \Xi)} \prod_{j=1}^{J} f(y_{jt}|X_t, \theta_d, \theta_c) \right). \end{aligned}$$

Therefore, the likelihood of the *observed data* is

$$
\begin{aligned}
&p(Y_{obs}, S, M|X, \theta_d, \theta_s, \theta_c)\\
&\qquad = \int p(Y, S, M|X, \theta_d, \theta_s, \theta_c) dY_{miss}\\
&\qquad = h(S|\theta_s, \theta_c) \prod_{t \in M_i} \left(\phi(1) \prod_{j=1}^{J} f(y_{jt}|X_t, \theta_d, \theta_c) \right)\\
&\qquad\quad \times \prod_{t \notin M_i} \left(\phi(1)^{I(t \notin \Xi)} \delta_0(m_t)^{I(t \in \Xi)} \prod_{j=1}^{J} \int f(y|X_t, \theta_d, \theta_c) dy \right)\\
&\qquad = h(S|\theta_s, \theta_c) \prod_{t \in M_i} \left(\phi(1) \prod_{j=1}^{J} f(y_{jt}|X_t, \theta_d, \theta_c) \right) \times \prod_{t \notin M_i} \phi(1)^{I(t \notin \Xi)} \delta_0(m_t)^{I(t \in \Xi)}\\
&\qquad = h(S|\theta_s, \theta_c) \left(\prod_{t \in M_i} \prod_{j=1}^{J} f(y_{jt}|X_t, \theta_d, \theta_c) \right) \times \prod_{t=1}^{T} \phi(m_t)^{I(t \notin \Xi)} \delta_0(m_t)^{I(t \in \Xi)}.
\end{aligned}
$$

Finally, considering again the individual indexes, the posterior distribution of the parameters is

$$
\begin{aligned}
p(\theta_d, \theta_s, \theta_c|Y, S, M, X) \propto\ & p(\theta_d, \theta_s, \theta_c) \prod_{i=1}^{N} h(S_i|\theta_s, \theta_c) \left(\prod_{t \in M_i} \prod_{j=1}^{J} f(y_{ijt}|X_t, \theta_d, \theta_c) \right)\\
& \times \prod_{t=1}^{T} \phi(m_{ti})^{I(t \notin \Xi)} \delta_0(m_{it})^{I(t \in \Xi)}\\
\propto\ & p(\theta_d, \theta_s, \theta_c) \prod_{i=1}^{N} h(S_i|\theta_s, \theta_c) \prod_{t \in M_i} \prod_{j=1}^{J} f(y_{ijt}|X_t, \theta_d, \theta_c). \qquad (6.14)
\end{aligned}
$$

Through a similar derivation we can also show that even if we consider that the disability response adopts a different state after death, say "*", by enlarging its support to be $\{0, 1\}^J \cup \{*\}$, we still reach a similar conclusion, provided that the three main assumptions are met. To do this all we have to do is to replace the model for y by

$$f^*(y_i|X_i,\theta) = \prod_{t=1}^{T} f(y_{i*t}|X_i,\theta)^{I(S_i>Age_{it})}\delta_*(y_{i*t})^{I(S_i\leq Age_{it})}.$$

Then proceed similarly considering separately the three cases: $Age_{it} < 65$ (ineligible), $Age_{it} \in [65, S_i]$ (possible to include in the sample), and $Age_{it} > S_i$ (dead).

Equation (6.14) shows that, under the stated assumptions, the joint model will take care of the missing data without further modeling of the process. This strong result is a consequence of the equally strong assumption of conditional independence between survival and disability, given membership scores.

6.4.5 Posterior Estimation Using MCMC

The augmented data representation shown in Equation (6.12) allows to easily expand the algorithm developed for the Basic Model in Chapter 2 (Section 2.2.1) to handle the survival distribution.

Considering uncensored survival data, the posterior distribution of parameters given the disability, survival and augmented data, and considering only the observed disability measurements, as suggested by the result in the previous section, is

$$\begin{aligned} p(\beta,\alpha, g, h|d, y, z, Age) \propto\ & p(\beta)p(\alpha_0)p\left(\xi\right)\left[\prod_{i=1}^{N} p\left(g_{i*}|\alpha\right)\right] \\ & \times \prod_{i=1}^{N}\left[\prod_{k=1}^{K}\left(g_{ik}h_k(d_i)\right)^{I(z_i'=k)} \times \prod_{j=1}^{J}\prod_{t\in M_i}\prod_{k=1}^{K}\left(g_{ik}\frac{\exp(y_{ijt}\beta_{0j|k}+y_{ijt}\beta_{1j|k}Age_{it})}{1+\exp(\beta_{0j|k}+\beta_{1j|k}Age_{it})}\right)^{I(z_{ijt}=k)}\right]. \end{aligned} \tag{6.15}$$

The expression of the posterior distribution in Equation (6.15) is similar to the one of the Basic model in Section 2. We can thus just adapt the MCMC sampler proposed there. We have to add steps to sample the parameters of the survival distribution and handle the data augmentation of the survival times and z'. It will also be necessary to modify the step for sampling g, as its full conditional distribution will have changed.

The algorithm can be implemented as follows:

1. **Sampling from z'_i (imputation)**

   $$z'_i|... \sim \text{Discrete}(p_1, ..., p_k),$$

   where

   $$p_k \propto g_{ik} \text{LTWeib}\left(d_i \left| \alpha_{dk}, \beta_{dk}, a_i \right.\right).$$

2. **Sampling from z_{ijt} (imputation)**: Same as the sampler for z_{ijt} in the Basic model.

3. **Sampling from g_i**: The full conditional distribution of g_i is

   $$p(g_i|...) \propto \prod_{j=1}^{J} \prod_{t \in M_i} \prod_{k=1}^{K} g_{ik}^{I(z_{ijt}=k)} \times \prod_{k=1}^{K} {g_{ik}}^{I(z'_i=k)}.$$

   Therefore, $g_i|... \sim \text{Dirichlet}(\alpha^g_{i1}, \alpha^g_{i2}, ..., \alpha^g_{iK})$, with

   $$\alpha^g_{ik} = \alpha_k + I(z'_i = k) + \sum_{j,t} I(z_{ijt} = k).$$

4. **Sampling from $(\beta_{0jk}, \beta_{1jk})$:** Same as the sampler for $(\beta_{0jk}, \beta_{1jk})$ in the Basic

model.

5. **Sampling from α:** Same as the sampler for α in the Basic model.

6. **Sample from d_i^* (Imputation)** This step is similar as the corresponding one in the sampler for the basic simple Weibull model in Section 6.3.1:

$$d_i^*|... = \begin{cases} d_i & \text{if } \nu_i = 0, \\ \text{Draw} \quad d_i^* \sim \text{LTWeib}(\alpha_{dz_i'}, \beta_{dz_i'}, b_i) & \text{if } \nu_i = 1, \end{cases}$$

where the censoring indicator, ν_i is defined to be $\nu_i = 0$ if the person was known to be alive at the time of the survival assessment and $\nu_i = 1$ otherwise.

7. **Sample from** $(\alpha_{dk}, \beta_{dk})$ The full conditional distribution is

$$\begin{aligned} p(\alpha_{dk}, \beta_{dk}|...) &\propto p(\alpha_{dk}, \beta_{dk}) \times \prod_{i=1}^{N} \left[\prod_{k=1}^{K} \text{LTWeib}\left(d_i^* \,|\alpha_{dk}, \beta_{dk}, a_i\,\right) \right]^{I(z_i'=k)} \\ &\propto \alpha_{dk}^{N_k + a_\alpha - 1} \beta_{dk}^{N_k \alpha_{dk} + a_\beta - 1} \\ &\quad \times \exp\left[\alpha_{dk} \left(\sum_{i \in \Xi_k} \log d_i^* - b_\alpha \right) + \beta_{dk}^{\alpha} \sum_{i \in \Xi_k} \left({a_i}^{\alpha_{dk}} - {x_i}^{\alpha_{dk}} \right) - b_\beta \beta_{dk} \right], \end{aligned}$$

where $\Xi_k = \{i \in 1...N : z_i' = k\}$ and $N_k = \#\Xi$.

Again, as in the algorithm in Section 6.3.1, we will use a Metropolis-Hastings step:

   (a) (proposal step) Sample a proposal value (α^*, β^*) from

$$\begin{aligned} \log \alpha^* &\sim N\left(\log \alpha_{dk}, \sigma_\alpha^2\right), \\ \log \beta^* &\sim N\left(\log \beta_{dk}, \sigma_\beta^2\right). \end{aligned}$$

(b) (Acceptance step) Compute

$$
\begin{aligned}
r' = &\ \frac{p(\alpha^*, \beta^*|....)}{p(\alpha_{dk}, \beta_{dk}|....)} \\
= &\left(\frac{\alpha^*}{\alpha_{dk}}\right)^{N_k+a_\alpha} \frac{\beta^{*\alpha^* N_k + a_\beta}}{\beta_{dk}^{\alpha_{dk} N_k + a_\beta}} \\
&\times \exp\Bigg[(\alpha^* - \alpha_{dk})\left(\sum_{i \in \Xi} \log d_i - b_\alpha\right) + \beta^{*\alpha^*}\left(\sum_{i \in \Xi}(a_i^{\alpha^*} - d_i^{\alpha^*})\right) \\
&\qquad - \beta_{dk}^{\alpha_{dk}}\left(\sum_{i \in \Xi}(a_i^{\alpha_{dk}} - d_i^{\alpha_{dk}})\right) - b_\beta\,(\beta^* - \beta_{dk})\Bigg],
\end{aligned}
$$

and update the vector $(\alpha_{dk}, \beta_{dk})$ from step (m) to $(m+1)$ by

$$
(\alpha_{dk}, \beta_{dk})^{(m+{}`1)} = \begin{cases} (\alpha^*, \beta^*) & \text{with probability } \min\{r', 1\}, \\ (\alpha_{dk}, \beta_{dk})^{(m)} & \text{with probability } 1 - \min\{r', 1\}. \end{cases}
$$

6.5 Summary and Discussion

In this chapter I have presented an extension to the Basic Grade of Membership Trajectory model aimed at joint model survival times and disability measurements. It is based on considering a submodel for the survival outcomes, consisting in individual-level mixtures of simple continuous distributions, and then linking it to a Basic GoM Trajectory submodel by allowing the two submodels to share the same membership vector. I also had to introduce corrections to the survival model in order to account for right censoring and delayed entry.

The resulting Joint model has the advantage of re-using the Basic model and thus retaining its interpretation and properties, and at the same time providing a way to account for the dependency of disability and mortality. Due to the way I have constructed the joint model, once estimated the parameters, what we will have is a characterization of a number K of extreme profiles that describe simultaneously typical tendencies in the aging process and survival for ideal individuals. At the same time, similar to the Basic model, the Joint model provides a characterization of the structure of the heterogeneity among real individuals, by means of the estimation of the distribution of the partial membership scores in the population.

This joint modeling is interesting for a number of reasons. First, the characterization and study of survival times is relevant in its own right, even more if it can shed light into the relationship between disability and mortality. Second, it allows to differentiate cases of missing data due to death from other cases. Finally, survival times are information that can complement and in some cases supplement disability measurements to allow for better estimates. This is especially important for cases where there are just a few measurements, as is the case with younger cohorts.

A relevant question related to the last point is what is the relationship between the estimates using the Basic model and the joint model? Using the general notation introduced in the discussion about the missing values, in Section 6.4.4, the likelihood of the model is

$$\Pr(Y = y, S = d | X, \theta_d, \theta_s, \theta_c) = f(y | X, \theta_d, \theta_c) \cdot h(d | \theta_s, \theta_c).$$

This means that the marginal likelihood of the joint model integrating over S is

$$\begin{aligned}\Pr(Y = y|X, \theta_d, \theta_s, \theta_c) =& \int \Pr(Y = y, S = x|X, \theta_d, \theta_s, \theta_c)dx \\ =& \int f(y|X, \theta_d, \theta_c) \cdot h(x|\theta_s, \theta_c)dx \\ =& f(y|X, \theta_d, \theta_c) \int h(x|\theta_s, \theta_c)dx \\ =& f(y|X, \theta_d, \theta_c),\end{aligned}$$

which is exactly that of the Basic model. Thus we see that the Basic model is the marginalized version of the joint model. Using the basic model will produce the same results that the joint model if we for some reason decide to average over the survival data.

Now let us see what happens when we perform inferences about the disability and common parameters under the Joint model using the whole dataset. The posterior distribution of all parameters given the data vector (Y, S, X) is

$$\begin{aligned}P(\theta_d, \theta_s, \theta_c|Y, S, X) \propto& P(\theta_d, \theta_s, \theta_c) f(Y|X, \theta_d, \theta_c) h(S|\theta_s, \theta_c) \\ =& P(\theta_d)P(\theta_s)P(\theta_c) f(Y|X, \theta_d, \theta_c) h(S|\theta_s, \theta_c),\end{aligned}$$

and therefore, the posterior distribution of the parameters that this Joint model have in common with the basic model is

$$\begin{aligned} P(\theta_d, \theta_c | Y, S, X) &\propto \int P(\theta_d, \theta_s, \theta_c | Y, S, X) d\theta_s \\ &\propto \int P(\theta_d) P(\theta_s) P(\theta_c) f(Y|X, \theta_d, \theta_c) h(S|\theta_s, \theta_c) d\theta_s \\ &= \underbrace{P(\theta_d) P(\theta_d) f(Y|X, \theta_d, \theta_c)}_{P(\theta_d, \theta_c | Y, X)} \underbrace{\int h(S|\theta_s, \theta_c) P(\theta_s) d\theta_s}_{w(S|\theta_c)} . \end{aligned}$$

This last expression is quite revealing: the inferences that we can make with the Joint model about the parameters (θ_c, θ_d)—those in common with the Basic model—will include information about survival. In fact the posterior distribution of those parameters is the one that we would have obtained from the application of the Basic model, $P(\theta_d, \theta_c | X, Y)$, modified by a term that depends on the survival information, $w(S|\theta_d)$.

This analysis makes explicit the mechanism through which data is shared between the survival and disability submodels. This sharing of data allows the use of to use survival outcomes to complement the disability information when the latter is scarce. In this way if, for instance, we only have one measurement for an individual, then its survival time, even if censored, can provide extra information to help the algorithms to better classify them with respect to the extreme profiles. Furthermore, this information also propagates to the estimates of the disability parameters themselves, borrowing strength from other similar individuals.

The way this effects the estimates of disability and in general the fit of the model, however, depends on the quality of the assumptions made to build the Joint model. In particular, I have assumed that the pure types can simultaneously describe survival and disability. This implies that, if patterns of disability and survival do not agree, the effect of the survival data in the disability estimates can potentially bias those estimates.

Chapter 7

Applying the Joint Disability-Mortality model to the NLTCS Data

In this chapter, I apply the joint disability-survival model from Section 6 to the NLTCS with the objective of testing the proposed methods. In particular I want to assess how the disability estimates change with respect to those produced by the Basic model alone; check how well the Joint model represents the survival outcomes and; to gain insight on the relationship between disability and mortality.

7.1 Preliminaries

The data for this analysis is the same sample from the NLTCS that was used with the two previous models. In addition, I have used the survival times from the CMS file that was originally used by Stallard (2005) and Connor (2006) in their analysis of the mortality in the NLTCS (see Section 1.1 in Chapter 1 for details). In terms of data preparation, I have re centered the ages at the time of death from the datafile to represent life in excess of 65, this is, substracted 65 to the age at the time of death. This way, I can take the support of the survival distributions to be the positive real line. I also created the indicator variable ν_i that takes the value 1 if the person was alive at the time of the assessment, in 2006, and 0 otherwise, as required for the Joint model.

For both the disability parameters, β_0 and β_1, and the GoM population-level parameters, α, I have used the same prior specification as with the Basic and Generational models: $\beta_{0jk} \overset{iid}{\sim} N(0, 100)$, $\beta_{1jk} \overset{iid}{\sim} N(0, 100)$, $\xi_k \sim \text{Dirichlet}(\mathbf{1}_K)$ and $\alpha_0 \sim \text{Gamma}(1, 5)$. The appropriateness of this choice follows from the same considerations presented in the Basic model case in Chapter 3. Additionally, as I want to make the estimates easily comparable, retaining the same prior specification will allow to me focus on the differences introduced by the model extension only.

For the survival prior distribution I have used the same parameters discussed for the test run of the simple Weibull model presented in Section 6.3.2, following the same considerations. These distributions are $\alpha_{dk} \overset{iid}{\sim} \text{Gamma}(1, 1)$ and $\beta_{dk} \overset{iid}{\sim} \text{Gamma}(1, 0.1)$ for all $k = 1, ..., K$.

Executing this algorithm is not very different from executing the other variants of the

Extreme Profiles	σ^*_α
$K = 2$	0.0650
$K = 3$	0.0116
$K = 4$	0.0085
$K = 5$	0.0070

Table 7.1: Values of proposal standard deviation σ^*_α for joint survival-disability model.

Latent Trajectories GoM model samplers. It presents the same challenges in terms of computing time and the need to fine-tune the proposal distribution for the Metropolis-Hastings steps. As in all the preceding cases, the proposal distribution most critically affected by fine tuning was the one for sampling the values of α. The final tuning parameters for these proposals, for models with $K = 1$ to $K = 5$ extreme profiles, σ^*_α, are shown in Table 7.1. For the survival tuning parameters, $\sigma^*_{d\alpha}$ and $\sigma^*_{d\beta}$, the values already chosen for the execution of the simple Weibull model (see Section 6.3.2), $\sigma^*_{d\alpha} = 0.01$ and $\sigma^*_{d\beta} = 0.2$ worked similarly well.

For all models, from $K = 2$ to $K = 5$, I have executed $100,000$ iterations, discarding the first $20,000$ samples as a burn in period and subsampled the rest, keeping one from every five samples and discarding the rest, to reduce the effect of the serial correlation of the chain.

7.2 Results

I have computed the posterior distribution of the parameters for models with $K = 2$ to $K = 5$ extreme profiles. Besides the population-level membership parameters, α_0 and ξ, the trajectory parameters, β_{0jk} and β_{1jk}, the survival parameters α_{dk} and β_{dk} and the summary $Age_{1/2}$, these tables also display the summary posterior $Age_{srv,k}$. This quantity is defined as the median survival age, in years since birth, of an ideal individual in profile k:

$$Age_{srv,k} = \frac{1}{\beta_{dk}} \left[\log 2\right]^{1/\alpha_{dk}} + 65. \tag{7.1}$$

I give posterior summaries of these parameters (posterior means and standard deviations) in Tables 7.4 to 7.6 in Appendix 7.A, at the end of this Chapter.

Figures 7.1 to 7.3 show the estimated trajectory functions for each combination of extreme profile and ADL, overlaid with the estimated profile's posterior survival function.

The first thing to note in these estimates is that they are similar to those obtained using the Basic model. They too define extreme profiles that, ordered decreasingly by the estimate of the component ξ_k, also reflect an increasing tendency to acquire disabilities earlier in life, as we consider profiles with lower and lower values of ξ_k.

The profiles themselves look quite sound, with survival curves that match the disability trajectories: profiles that imply a late onset of disability (e.g., $k = 1$, for model with $K = 3$) also have associated survival curves with long expected survival times, while profiles that reflect early frailty (e.g., $k = 3$ for model with $K = 3$) imply shorter survival times. The estimates of α_{dk} are all greater than 1, for all models, showing increasing hazards as subjects age.

The population-level membership vector distribution parameters, ξ and α_0 are also very similar to the ones computed using the Basic Model, although the Joint model tends to put more weight (greater ξ_k) into profiles with late onset of disability, while at the same time it increases the mixing of extreme profiles—expressed through a larger value of α_0. This effect is likely due to the added complexity of the response, that now includes the survival in addition to the disability.

Comparing the trajectories of disability obtained with the Joint model with those obtained from the Basic models, (see Figure 7.4, for a comparison with models with $K = 3$), we see that all trajectories from the Joint model are shifted to the left. We can reach the same conclusion by comparing the corresponding posterior summaries $Age_{1/2,jk}$ (see tables 7.4 to 7.6 in Appendix 7.A) which, under the Joint model, are all smaller than their corresponding ones under the Basic model.

While these differences may have been expected, given that the posterior estimates in the Joint model are, even for disability responses, including information from the survival response, it is noteworthy that these differences are so systematic. To understand this, we have to consider that while ADLs are merely indicators of a latent process, death is a clear a non reversible event and therefore introduces high quality information into the estimation process. Nonetheless, we still have to evaluate whether these changes could be considered an improvement or not. This can be assessed by the analysis of the predictive qualities of the Joint model.

Table 7.2 presents posterior predictive counts of the disability response, computed in a similar way as Tables 3.6 and 5.6 in their respective Chapters. The comparison with the basic model is quite instructive. Using the X^2 statistic computed from the contingency tables as a comparison reference (see Table 7.3 for a quick comparison table), we can see that the Joint model has in general a disability fit comparable with a Basic model with one extra extreme profile. Thus the fit of the Joint model to disability responses data can be considered as a remarkable improvement on the fit obtained by means of the Basic model alone.

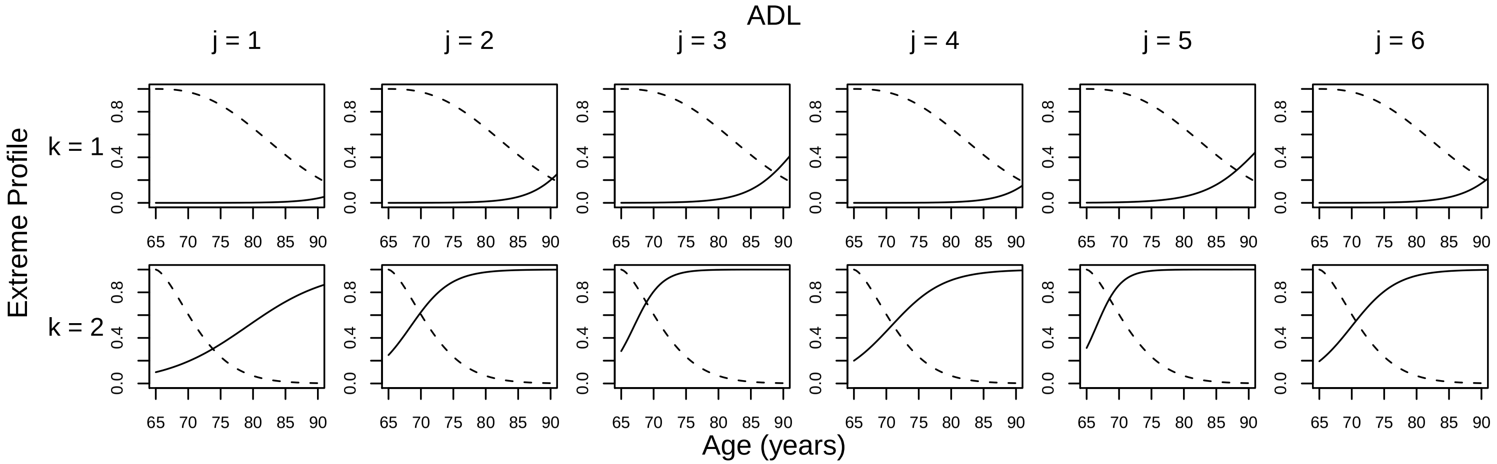


Figure 7.1: Extreme profiles of survival probability and disability trajectories ($K = 2$).

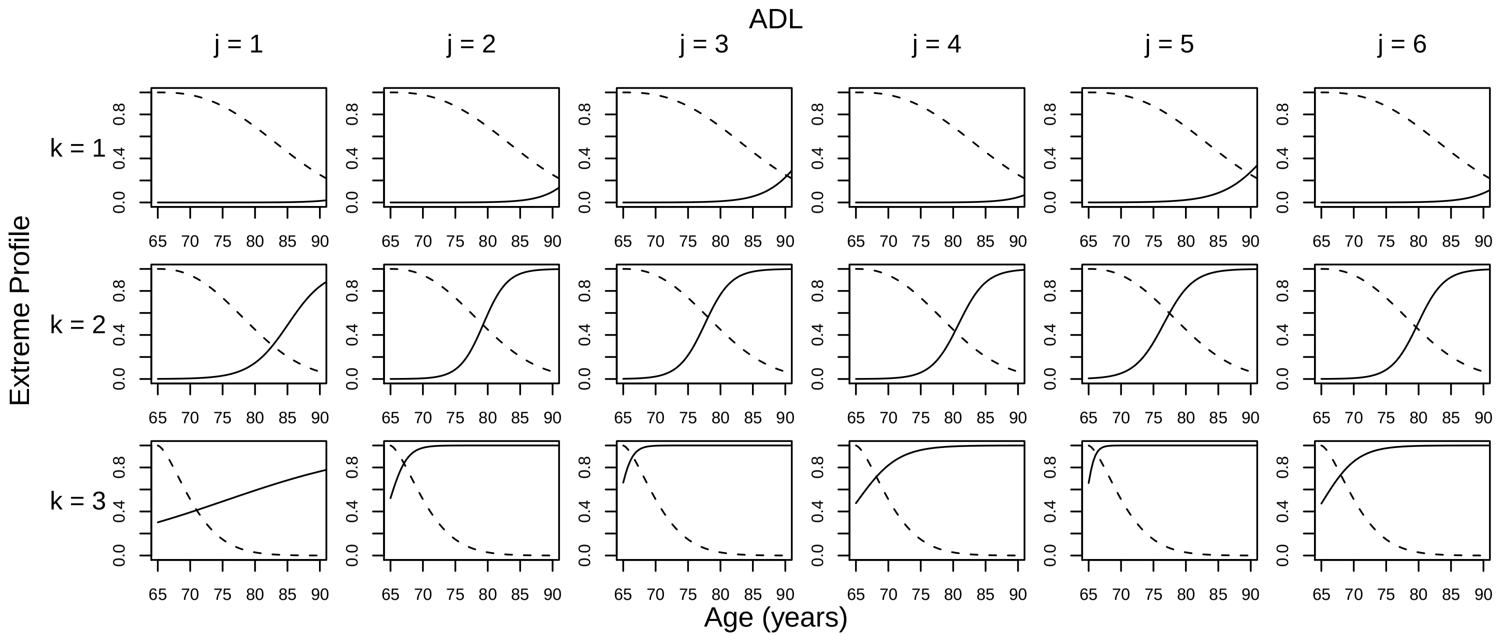


Figure 7.2: Extreme profiles of survival probability and disability trajectories ($K = 3$).

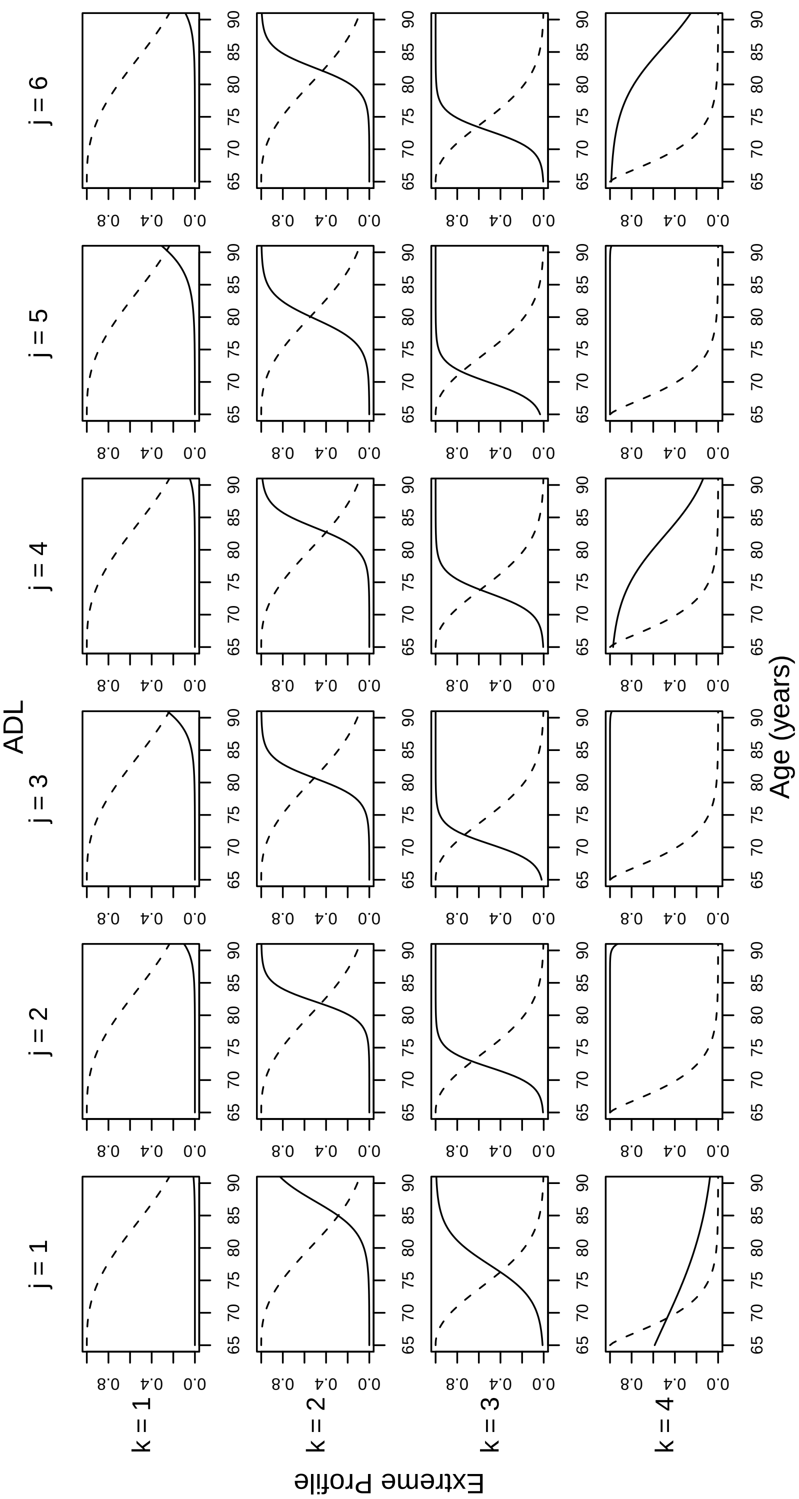


Figure 7.3: Extreme profiles of survival probability and disability trajectories ($K = 4$).

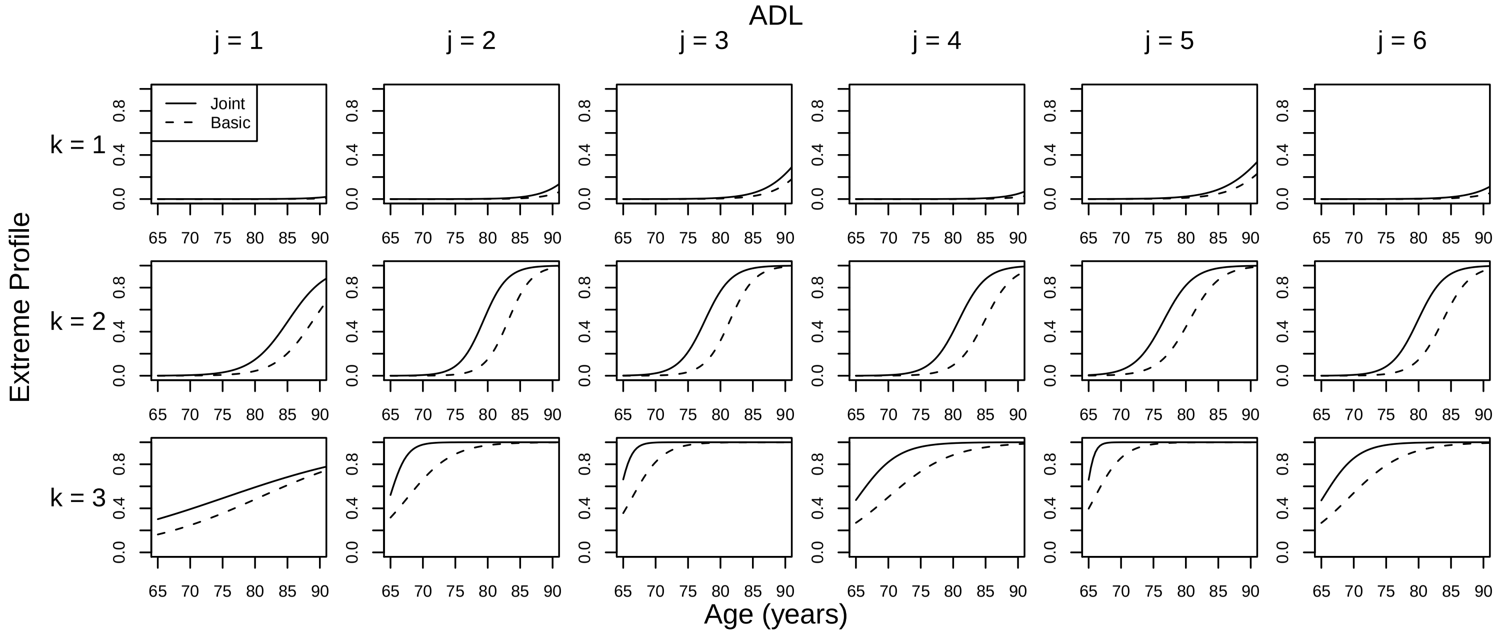


Figure 7.4: Comparison between disability extreme trajectories from Joint (solid line) and Basic (dashed line) models with $K = 3$ extreme profiles.

			Extreme Profiles		
Age Range	Pattern	Observed	$K = 2$	$K = 3$	$K = 4$
$65 - 70$	000000	21361	20128	21143	21196
	111111	295	4	89	217
	010000	292	475	256	324
	111110	192	25	184	257
	000100	167	366	161	214
	010110	117	96	97	76
	010100	103	170	86	87
	000110	89	93	65	65
$70 - 79$	000000	34508	32836	32800	33513
	010000	1143	1317	1533	1382
	111111	1087	377	527	559
	111110	692	713	732	820
	000100	584	772	931	746
	010100	412	338	483	403
	010110	375	316	271	262
	110110	335	386	249	316
$79 - 80$	000000	15536	12695	13911	14073
	111111	1734	934	950	1006
	010000	1178	2115	1535	1396
	111110	1018	789	1004	923
	000100	562	1486	1012	894
	010100	536	489	390	497
	010110	529	242	320	394
	110110	461	241	414	443
$90 <$	000000	1311	258	478	595
	111111	1025	386	574	682
	111110	529	276	281	313
	010110	196	176	152	145
	010000	193	240	318	320
	010100	189	239	244	229
	110110	158	167	137	133
	000100	138	217	271	267
		X^2	36562	9227	6288

Table 7.2: Observed and mean predictive posterior aggregated counts for the 10 most populated response patterns by age range, for basic model with $K = 2, 3, 4, 5$. Within each age group the response patterns are sorted decreasingly according to observed frequencies.

Model	$K = 2$	$K = 3$	$K = 4$
Basic	141,970	15,343	4,638
Joint	36,562	9,272	6,288

Table 7.3: Comparison between fit summary (X^2 statistic computed from posterior predictive disability response counts stratified by age categories) of Basic and Joint model for different values of K.

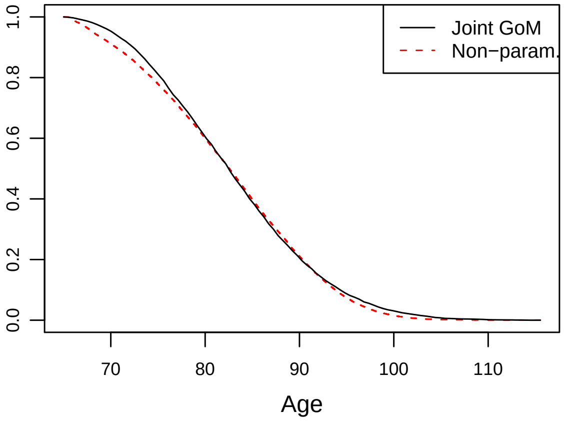


Figure 7.5: Estimated population-level survival curve for the joint model with $K = 3$ extreme profiles, compared with a non-parametric hazards based estimate as reference.

These results show that the inclusion of the survival information is in fact helping to obtain better predictions of disability, as I had conjectured in Section 6. This effect is not at all obvious. We have to bear in mind that the Joint model itself is forcing the membership vector to explain not only disability but also survival heterogeneity. This effect also shows that the hypothesized relationship between disability and survival, with both being considered as expressions of a same underlying process summarized by the membership vector, is at least fruitful in terms of allowing to share information between these phenomena.

To give an idea of the survival prediction fit, Figure 7.5 shows the estimated population-level posterior survival distribution, overlaid for comparison with a nonparametric survival estimate with censoring and delayed entry correction, similar to the one in Figure 6.2 in Chapter 6. Visually we can appreciate that the posterior curve of the joint model is indeed very close to the nonparametric estimate (which does not use the disability information). We can also note that it behaves better than the Weibull model from Section 6.3.2 near the origin, likely due to the increased flexibility of the representation, by the introduction of the mixed membership apparatus.

7.3 Evaluating the Survival Truncation Approximation

When we discussed the construction of the Joint Model in the previous chapter, it became clear that using the true form of the left truncated version of the survival response under the GoM Weibull model from Section 6.4.1,

$$p(d_i|g_i, a_i) \quad = \quad \frac{\sum_{k=1}^{K} g_{ik} \text{Weibull}(d_i|\alpha_k, \beta_k)}{\sum_{k=1}^{K} g_{ik} \left(1 - \int_0^{a_i} \text{Weibull}(x|\alpha_k, \beta_k) dx\right)}, \tag{7.2}$$

needed to perform the delayed entry correction would make the final expression too complex for estimation. Then I suggested to replace it with the expression

$$p'(d_i|g_i, a_i) \quad = \quad \sum_{k=1}^{K} g_{ik} \frac{\text{Weibull}(d_i|\alpha_k, \beta_k)}{1 - \int_0^{a_i} \text{Weibull}(x|\alpha_k, \beta_k)dx}, \qquad (7.3)$$

that corresponds to the operation of truncating each component individually and then performing the mixing of components using the individual membership vector, g_i. These two expressions are in general different, but they become equal if the membership vector expresses full membership in any of the extreme profiles (i.e. $g_k = 1$ for some k). Thus we can expect them to be approximately equal if the membership vectors tend to be closer to one particular component. This seems to be the case, given the small estimates of the population-level membership concentration parameter, α_0.

Figure 7.6 shows a comparison between the densities evaluated using the true expression (Equation 7.2) and the approximation (Equation 7.3). These values were computed using the posterior mean estimates of α_{dk} and β_{dk} and, from a sample of 1,000 individuals, the entry ages a_i and posterior predictive survival times, d_i ($K = 3$). In this way the most frequent cases should be represented. The comparison confirms the conjecture. Most of the points fall into the diagonal line, showing that for those individuals the approximation is working quite well. There are some points for which the discrepancy can be higher (although not extreme), and in general the approximation tends to overestimate the true expression. However, this occurs in just a small fraction of cases.

This evidence is of course not conclusive and has to be evaluated in combination with the rest of the results. While it certainly could be the case that the use of the approximation is already driving the algorithm to move along regions in the parameter space that make

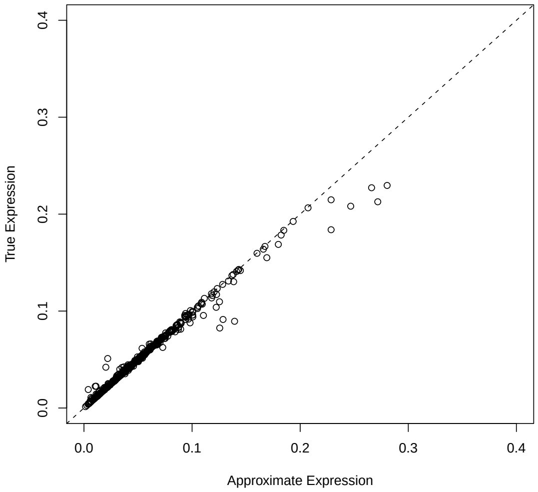


Figure 7.6: True value of the truncated mixed membership survival density (Equation 6.9) evaluated at the posterior expected time of death Vs. surrogate approximation (Equation 6.10) for 1000 random individuals from the sample. Model fitted with $K = 3$ extreme profiles.

the discussed comparison look better than it should, we must also bear in mind that 1) the estimates are quite sound and in terms of prediction are in fact an improvement over the previous models and 2) the estimated survival curve is perfectly comparable to the one computed using a nonparametric approach that also handles left truncation.

7.4 Understanding the Trajectory Profiles: Using Step Trajectories with the Joint Model

As we discussed in Chapter 3, in Section 3.3, we can gain some insight into the underlying tendencies of aging by using a less constrained specification for the extreme trajectories. To this end, I have implemented a version of the Joint model using piecewise constant ("step") functions, similar to the one proposed in Chapter 2.3 and applied in Chapter 3.3. I have used almost the exact specification as in the discrete trajectories variant of the Basic model, in Chapter 3, in particular the same breakpoints, $c_1 = 68, c_2 = 71, c_3 = 74, c_4 = 77, c_5 = 80, c_6 = 83, c_7 = 86$, and flat priors for the steps of the extreme trajectories. The specification for the survival distribution was exactly the same as with the Joint model, using the GoM Weibull model with right-censoring and left-truncation correction. I estimated all the parameters of the model simultaneously.

Figure 7.7 shows a graphical representation of the estimated trajectory functions for each extreme profile, for a model with $K = 3$. The trajectories are superposed with the estimated survival curve of each corresponding extreme profile. The rest of the estimates are very similar to those obtained with the Joint model with continuous trajectories.

One thing that we observed when applying the discrete trajectories to the NLTCS data using the Basic model, was that the continuous and discrete trajectories were very similar, except in the less predominant profiles, where in both versions trajectories tended to behave erratically. Interestingly enough, the situation with the Joint model is similar. Discrete step trajectories are very similar to the continuous ones, except in the last extreme profile $k = 3$, where starting around age 83, they exhibit a behavior that can be qualified as erratic.

The inspection of the mortality characterization of profile $k = 3$ seems to confirm the conjecture introduced in Section 3.3: less dominant profiles, that are also profiles with earlier onset of disability, are also characterized by earlier mortality. This means that individuals that are mostly described by traits from those profiles also tend to die earlier, providing little data to estimate disability at late ages. In Figure 7.7 this situation is evident. According to the estimates, the probability of surviving past age 83 for members of the $k = 3$ extreme profile is just $\Pr(S > 83|g_{i3} = 1) = 0.00922$, while the probability of the same event for members of profiles $k = 1$ and $k = 2$ is 0.557 and 0.287, respectively.

From a practical perspective, the results from this analysis suggest that the specification of strongly parametrized trajectories, like the monotone continuous curves proposed by Connor (2006) and also applied here, actually helps to obtain better estimates. These specifications have the net effect of using the data from regions where it is abundant to estimate the corresponding parameters, and then to extrapolate those tendencies into regions with a scarcity of data points. The similarities of the trajectories estimated in this section (step functions; Figure 7.7) and those obtained through the use of continuous parametric curves provide a good indication that the chosen parametric curves are an adequate choice.

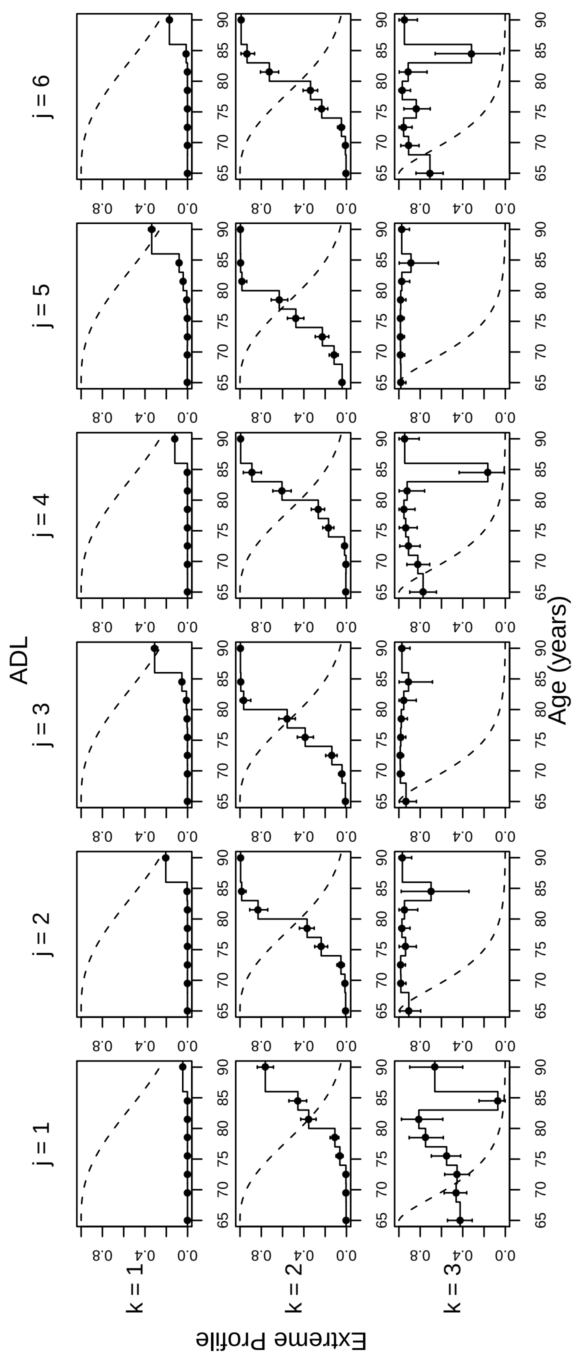


Figure 7.7: Extreme trajectories of the joint model with discrete trajectories specification.

7.5 Discussion

The application of the joint disability-survival models introduced in Chapter 6 to the NLTCS data produced results that are interesting in several accounts. In the first place, they provide a better picture of the main tendencies in the population, by complementing the extreme profiles with descriptions of mortality. This way, the inspection of those joint profiles shows that in the NLTCS, patterns of mortality are correlated with patterns of disability in a regular way: profiles that express "healthy aging" go together with long survival times, while profiles that express earlier increases in the probabilities of disability go together with higher probabilities of death at younger ages.

The way I have set up the models, with the survival outcomes as an extension of the response vector that also include the longitudinal disability outcomes, adds some interesting features to this analysis. As I had discussed in Chapter 6, the inclusion of the survival response was expected to affect the disability estimates, but not necessarily for good. By postulating a joint dependence between disability and mortality, these models are effectively sharing data between these two phenomena to estimate a common pool of parameters. I had hypothesized that, if the relationship between disability and survival was adequately described by the model, the joint estimation was going to improve both the disability estimates and the mixed membership classification.

This was indeed the case. Comparing the predictive capabilities of the Joint model to those of the Basic model I have found that, while in general terms they produce similar estimates, the Joint model produced better posterior predictive estimates of the disability outcomes. This means that survival data can in fact be used to complement disability data, helping to take care of missing data due to death and allowing to differentiate individuals for whom the disability information is scarce.

As I showed in Chapter 6, Section 6.5, survival data also propagates to the estimates of the disability parameters. This was verified empirically as a shift to the left of all estimated extreme trajectories of disability, with respect to those estimated by means of the Basic model alone. This effect was almost exactly the same for all extreme trajectories, in all the fitted models. This fact, together with the improvement achieved in predictive capabilities, suggests that the inclusion of survival information is helping to reduce a bias of the Basic model, likely to be consequence of the non-ignorability of missing data due to death (see Section 6.4.4 in Chapter 6 for details).

7.A Appendix - Posterior Summaries

Parameter	Estimate
α_0	0.599 (0.016)

Parameter	ADL(j)	Estimate Extreme Profile-k (sd)			
		$k = 1$		$k = 2$	
ξ	–	0.880	(0.002)	0.120	(0.002)
α_d	–	2.509	(0.013)	1.533	(0.061)
β_d	–	0.047	(0.000)	0.127	(0.007)
β_{0**}	1 (EAT)	-6.120	(0.100)	0.146	(0.037)
	2 (BED)	-4.393	(0.052)	3.748	(0.157)
	3 (MOB)	-3.411	(0.035)	6.148	(0.296)
	4 (DRS)	-5.057	(0.068)	2.269	(0.086)
	5 (BTH)	-2.859	(0.027)	7.134	(0.386)
	6 (TLT)	-4.338	(0.050)	2.871	(0.113)
β_{1**}	1 (EAT)	0.295	(0.007)	0.157	(0.005)
	2 (BED)	0.301	(0.005)	0.323	(0.013)
	3 (MOB)	0.277	(0.004)	0.472	(0.023)
	4 (DRS)	0.303	(0.005)	0.244	(0.008)
	5 (BTH)	0.239	(0.003)	0.528	(0.029)
	6 (TLT)	0.276	(0.004)	0.286	(0.010)
$Age_{1/2}$	1 (EAT)	100.747	(0.230)	79.073	(0.224)
	2 (BED)	94.618	(0.119)	68.409	(0.165)
	3 (MOB)	92.308	(0.099)	66.960	(0.135)
	4 (DRS)	96.671	(0.142)	70.691	(0.181)
	5 (BTH)	91.958	(0.104)	66.498	(0.131)
	6 (TLT)	95.726	(0.138)	69.963	(0.164)
Age_{srv}	–	83.299	(0.000)	71.177	(0.007)

Table 7.4: Posterior means for parameters of interest for survival model with $K = 2$ extreme profiles. Numbers between parenthesis are posterior standard deviations.

Parameter	Estimate
α_0	0.385 (0.011)

Parameter	ADL(j)	Estimate Extreme Profile-k (sd)					
		$k = 1$		$k = 2$		$k = 3$	
ξ	–	0.751	(0.007)	0.194	(0.006)	0.055	(0.001)
α_d	–	2.550	(0.017)	2.383	(0.046)	1.515	(0.063)
β_d	–	0.045	(0.000)	0.061	(0.001)	0.154	(0.008)
β_{0**}	1 (EAT)	-7.688	(0.204)	-1.760	(0.061)	0.369	(0.086)
	2 (BED)	-5.832	(0.109)	0.363	(0.092)	11.302	(1.389)
	3 (MOB)	-4.469	(0.071)	1.195	(0.093)	16.407	(1.926)
	4 (DRS)	-6.646	(0.137)	-0.402	(0.079)	4.745	(0.400)
	5 (BTH)	-3.747	(0.058)	1.515	(0.089)	24.326	(4.065)
	6 (TLT)	-5.605	(0.101)	0.064	(0.081)	5.533	(0.435)
β_{1**}	1 (EAT)	0.345	(0.011)	0.343	(0.009)	0.081	(0.010)
	2 (BED)	0.362	(0.007)	0.542	(0.015)	0.748	(0.102)
	3 (MOB)	0.325	(0.005)	0.497	(0.013)	1.049	(0.140)
	4 (DRS)	0.366	(0.009)	0.474	(0.013)	0.323	(0.032)
	5 (BTH)	0.279	(0.005)	0.442	(0.010)	1.578	(0.287)
	6 (TLT)	0.322	(0.007)	0.490	(0.013)	0.376	(0.034)
$Age_{1/2}$	1 (EAT)	102.313	(0.301)	85.130	(0.199)	75.461	(0.670)
	2 (BED)	96.121	(0.142)	79.332	(0.163)	64.858	(0.307)
	3 (MOB)	93.748	(0.124)	77.595	(0.165)	64.320	(0.401)
	4 (DRS)	98.181	(0.175)	80.850	(0.175)	65.283	(0.383)
	5 (BTH)	93.415	(0.131)	76.574	(0.169)	64.536	(0.365)
	6 (TLT)	97.397	(0.174)	79.872	(0.165)	65.278	(0.337)
Age_{srv}	–	84.095	(0.000)	79.117	(0.001)	70.103	(0.008)

Table 7.5: Posterior means for parameters of interest for survival model with $K = 3$ extreme profiles. Numbers between parenthesis are posterior standard deviations.

Parameter	Estimate
α_0	0.326 (0.010)

Parameter	ADL(j)	Estimate Extreme Profile-k (sd)							
		$k=1$		$k=2$		$k=3$		$k=4$	
ξ	–	0.670	(0.011)	0.226	(0.010)	0.071	(0.002)	0.033	(0.001)
α_d	–	2.615	(0.024)	2.551	(0.048)	2.209	(0.072)	1.346	(0.092)
β_d	–	0.044	(0.000)	0.055	(0.001)	0.085	(0.002)	0.194	(0.018)
β_{0**}	1 (EAT)	-8.871	(0.284)	-2.950	(0.078)	0.961	(0.106)	-1.305	(0.109)
	2 (BED)	-7.288	(0.181)	-1.482	(0.092)	5.839	(0.290)	20.065	(4.876)
	3 (MOB)	-5.431	(0.111)	-0.404	(0.086)	6.771	(0.284)	23.056	(4.646)
	4 (DRS)	-8.075	(0.209)	-2.026	(0.085)	4.324	(0.224)	0.431	(0.153)
	5 (BTH)	-4.459	(0.087)	0.145	(0.080)	6.941	(0.357)	24.417	(5.461)
	6 (TLT)	-6.629	(0.145)	-1.556	(0.086)	4.979	(0.264)	1.261	(0.261)
β_{1**}	1 (EAT)	0.416	(0.016)	0.411	(0.011)	0.367	(0.016)	-0.110	(0.010)
	2 (BED)	0.464	(0.012)	0.706	(0.023)	0.726	(0.033)	-1.587	(0.377)
	3 (MOB)	0.397	(0.009)	0.631	(0.018)	0.714	(0.029)	-1.726	(0.351)
	4 (DRS)	0.461	(0.013)	0.595	(0.019)	0.645	(0.027)	-0.206	(0.020)
	5 (BTH)	0.332	(0.007)	0.518	(0.015)	0.688	(0.033)	-1.829	(0.398)
	6 (TLT)	0.391	(0.010)	0.616	(0.019)	0.692	(0.030)	-0.214	(0.029)
$Age_{1/2}$	1 (EAT)	101.332	(0.291)	87.187	(0.188)	77.386	(0.211)	68.151	(0.681)
	2 (BED)	95.721	(0.144)	82.100	(0.142)	71.963	(0.145)	92.635	(0.625)
	3 (MOB)	93.677	(0.136)	80.641	(0.138)	70.512	(0.145)	93.378	(0.656)
	4 (DRS)	97.511	(0.175)	83.409	(0.160)	73.295	(0.152)	82.108	(0.762)
	5 (BTH)	93.437	(0.145)	79.722	(0.151)	69.908	(0.137)	93.341	(0.566)
	6 (TLT)	96.957	(0.179)	82.527	(0.148)	72.802	(0.144)	85.900	(0.859)
Age_{srv}	–	84.619	(0.000)	80.875	(0.001)	74.957	(0.002)	68.925	(0.018)

Table 7.6: Posterior means for parameters of interest for survival model with $K = 4$ extreme profiles. Numbers between parenthesis are posterior standard deviations.

Chapter 8

Summary and Future Work

In this thesis I introduce a new family of models, the Grade of Membership Trajectory models, that are specially well suited to characterize and analyze typical multivariate trends of binary—or categorical by extension—responses over time, while at the same time allowing for complex forms of individual heterogeneity in the population. These models handle the individual heterogeneity using the concept of partial membership: assuming the existence of a small number of ideal types or extreme profiles and letting each individual to partially belong to each pure type, in different degrees. To model the general trends over time, these models start by characterizing the ideal individuals (exclusive members of the extreme profiles), using "extreme trajectories": curves that describe the underlying probabilities of having positive outcomes in each component of the response as a function of time, for each extreme profile. Then, the description of the real progressions over time for the non-ideal (or real) individuals is constructed by combining those extreme trajectories, weighting them by the degree of membership of the individual within the corresponding profile. This way, each individual can follow a different trajectory over time, but this trajectory can be described as a combination of the population-level defined extreme profiles.

This development was motivated by an interest in analyzing disability patterns in elder Americans using the NLTCS. This way, I used the extreme trajectories to represent the evolution of the extreme tendencies in the probability of acquiring specific disabilities as the individual ages. I used the mixed membership apparatus to characterize the population heterogeneity, looking for flexibility to allow for different and personal ways of aging.

8.1 Contributions

The development of these models and the associated methods for estimation are a contribution to the current literature on Mixed Membership models, in particular the Grade of Membership model (Woodbury et al., 1978; Manton et al., 1994; Erosheva et al., 2007), extending this family by introducing covariates into the extreme profile characterizations. At the same time, they can be considered as another layer of sophistication on the Multivariate Latent Trajectory models (Connor, 2006), replacing the exclusive clustering of the population with a more flexible soft-clustering, using mixed membership ideas.

I have also developed two main extensions to this basic idea, both aimed at providing insights into the general topic of disability of elderly people in the U.S. using the available data from the NLTCS. The first one directed at answering questions about the differences in the ways of aging across different generations ("Do younger generations acquire disabilities in a different way than older ones?"). I framed the problem extending the models, introducing cohort information (the generational group of each individual) as a covariate that affected the membership distribution. This way, differences between generations could be read as differences in the way that the cohort information affected the membership distribution, using the extreme profiles as a common reference frame. This approach is an instance of the more general idea of including covariate information into the population-level membership distribution, and it can be used to handle other stratifications of the population, including those introduced by unequal probability sampling.

The second extension incorporates information about survival times, combining it with disability outcomes. It is particularly relevant for analyzing data from the NLTCS because disability and mortality in advanced ages are phenomena that are interrelated and the joint modeling helps to understand the relationship between them. In technical terms, the extension is interesting for several reasons: it combines categorical and continuous outcomes into a complex multidimensional response vector, using a common mixed membership framework; it shows how to incorporate nontrivial sampling issues related (but not limited) to the characteristics of survival-time data, like delayed entry (left truncation) and right censoring; it presents a simple way for accounting for missing values due to death and; it shows how we can use complementary information (survival data) to enhance the estimation of both the mixed membership distribution and the parameters that directly model the disability outcomes, through simultaneous inference.

In terms of estimation, I have developed a series of algorithms based on Markov Chain Monte Carlo sampling that extend the one introduced in Erosheva (2002) and Erosheva et al. (2007). These algorithms are based on an equivalent Latent Class representation of the GoM model first proposed in Haberman (1995) and fully developed in Erosheva (2002) and Erosheva et al. (2007). I have developed algorithms and variants for each of the models proposed and showed them in action by fitting the models to the NLTCS data.

In terms of the motivating problem, this work contributes a new set of tools and methods to the still small body of research about disability of elderly people in the United States through the use of longitudinal methods applied to the NLTCS data (Stallard, 2005; Connor, 2006; White, 2008; Manton et al., 2008). Different from other studies in the past, that have relied on the analysis of the waves of the NLTCS as a set of uncorrelated samples (see, e.g., Manton et al. 1997, 2006, 2007), the methods I introduce here use the longitudinal data structure present in the NLTCS, consisting of sequential measurements of the same individuals over time. This added complexity enables the data analysts to draw conclusions at the level of individual change over time.

Applying the Basic GoM Trajectory model to the NLTCS data (Chapter 3), I have shown how general tendencies of the aging of Americans can be effectively modeled as monotone increasing extreme trajectories and how the general population can be characterized by combinations of these profiles. The introduction of the mixed membership apparatus helps us to keep a relatively small number of extreme profiles while still allowing the representation of complex trajectories. I have also shown that most individuals tend to be closer to profiles that describe a late onset of disability and that, as we consider profiles that describe an earlier onset of disability, they are less predominant in the general population.

Using the Grouped data extension (Chapter 5), I have been able to provide evidence supporting a decline in disability in elder Americans over time: I have shown that, as we consider younger generations, the individuals that belong to these tend to be closer and closer to a profile that describes a very late onset of disability. This analysis separates the effect of the age of the individuals from their dates of birth, focusing on their ways of aging, conditional on their cohort, rather than on pure populational prevalences. Prevalences can also be estimated integrating the estimated conditional distribution of membership with respect to the distribution of cohorts in the population, using supplemental data from, for instance, the Census or Medicare records.

It is worth noting that this cohort analysis provides, from an alternative perspective to existing studies, new evidence supporting a general decline in disability in U.S. elders over time. Decline of disability prevalence was already observed from cross-sectional (uncorrelated) estimates (Manton et al., 1997, 2006) and wave to wave latent class transitions (White, 2008). Connor (2006) also presents results pointing inthe same direction, from a similar perspective to mine, using trajectory estimates.

The application of the Joint survival-disability model extension (Chapter 7) shows that the inclusion of the mortality information has the effect of improving the quality of the estimates of disability. This fact alone would be enough justification for the added complication of performing the joint estimation. Additionally, this analysis shows that profiles that express an early onset of disability are also characterized by higher death probabilities. Although I have limited the analysis to the obvious features of the combined disability-survival profiles and the distribution of the population with respect to those profiles, the proposed method actually reconstructs a joint probability distribution of disability and survival. These estimates can potentially provide answers to more complex public policy questions, like conditional probabilities of death given particular disability combinations.

8.2 Directions for Future Work

1. **Combining the Grouped and Survival extensions**:

After developing two extensions to the Basic GoM Trajectory model to account for two independently relevant aspects on the study of disability, the obvious next step is to combine them together into a general model that takes those two aspects into consideration simultaneously. Such an extension would actually be very simple to create, based on the work already done on the Grouped (Chapter 4) and Joint Survival-Disability (Chapter 6) models. As these extensions affect relatively independent parts of the Basic model, it is possible to derive an MCMC sampler for a combined model basically reusing and combining the already proposed steps for those two variants. A model like this would more accurately determine the extreme trajectories and group membership, by borrowing strength from the survival outcomes and therefore would provide more reliable estimates.

2. **Including other Covariates**

   Some authors (see e.g. White 2008; Manton 2008) have pointed out that certain health-related events that evolve over time, for instance acquiring Alzheimer's decease or arthritis, are strong predictors of disability. The NLTCS includes measurements of several of these potentially relevant events. Although it can be argued that events whose probability increases with age, like the ones mentioned, can be implicitly accounted for in the extreme trajectory definitions, it seems fruitful to explore forms of accounting for them explicitly, either as covariates or as responses. This can help obtain better estimates, like in the case of the simultaneous modeling of survival and disability, and/or to study the interaction between these factors and disability.

   There are other potentially relevant static characteristics of the individuals that could also be fruitfully incorporated into the models. An obvious candidate is sex, as it has been shown repeatedly that men and women follow different aging and mortality processes (see e.g. Ferrucci et al. 1996; Connor 2006; Manton et al. 2008). One way of incorporating these covariates is to directly use the Grouped model (Section 4), partitioning the population according to the cartesian product of the discrete (or discretized) covariates.

3. **Including the IADLs**

   All the analyses performed using data from the NLTCS in this thesis have been done using only the six ADLs. Extending the models and estimation techniques to account for the IADLs should, in principle, be a straightforward exercise if we consider ADLs and IADLs exchangeable. However, even under this simplifying assumption, since no IADL measurements are registered for institutionalized individuals, we still must account for the missing data somehow. One way of doing it is to model explicitly the censoring process and use a data-augmentation algorithm, such as the ones in (Tanner, 1996), for estimation.

4. **Period analysis**

   Although I have intentionally not considered period effects into my analyses, choosing instead to explain all temporal variability as a combination of age and cohort effects (see Chapter 4), there are good reasons to consider at least some form of period effects. Medical breakthroughs and changes in public policy affect all individuals simultaneously—although not necessarily in the same way—regardless of their age or cohort (Manton et al., 2008). Additionally, important public policy questions can only be answered satisfactorily by some form of period analysis, specially when they refer to the evaluation of the effectivity of interventions or changes in public policy.

5. **Model Selection:**

   The problem of determining the optimal number of components (K) is common to all latent structure models. The models presented in this thesis are no exception and the problem of choosing the "right" number of extreme profiles remains open. In my analyses I have addressed these issues in a rather informal way, by comparing the models' fit using summaries of posterior predictive responses and qualitative considerations in the context of the problem. While I do believe that for this particular problem this approach provides a good insight into the problem, a more systematic way of performing this selection would be desirable. Some alternatives that have been used in the past with Mixed Membership models, although not in longitudinal settings, include the evaluation indexes that balance goodness of fit and complexity such as BIC, AIC, or DIC (Erosheva et al., 2007).

   Other alternatives, from a purely Bayesian standpoint, include the estimation of posterior model distributions through the use of algorithms such as the Reversible Jump Markov Chain Monte Carlo (Green, 1995). Additionally, non-parametric techniques, such as the Dirichlet Process priors (Ferguson, 1973; Airoldi et al., 2007) can also be considered.

6. **Other computing strategies:**

   In this thesis I have developed estimation techniques based on Markov Chain Monte Carlo sampling. These methods have many advantages including computational tractability and the generation of estimates for the whole joint posterior distribution, instead of simple point estimates. They can, however, require huge amount of computing power, as each iteration of the Basic model is required to generate and keep track of a latent space of more than $N \times T \times J$ parameters (in the order of

$1,260,000$, in this application). This can be very slow, taking in the order of days to complete the computations for each single model in this thesis. While in this application computing times are not really critical, as we are not faced with a continuous flow of new data, computing power requirements limit their scalability. For these models to be useful in other applications, faster and less resource-consuming algorithms would be desirable. Estimation algorithms based on the idea of mean field variational inference have been applied successfully on similar problems, although not longitudinal (Airoldi et al., 2008; Blei and Lafferty, 2007; Erosheva et al., 2007), and could be adapted to work in this framework.

7. **Longitudinal Dependence:**

   My models do not explicitly account for serial correlation. Although some form of temporal correlation is accounted for by the membership scores, specifically through the assumption of conditional independence given membership scores, we have to bear in mind that those same membership scores are also being used to model the dependence between response components (the different ADLs). Thus explicit modeling of the serial correlation, by different means, would be desirable.

   Additionally, when posed in full generality, the problem of the serial correlation in this context can be extremely complicated, given the continuous nature of the aging process, opposed to the discrete nature of measurements through survey waves. One question to be answered is if the serial dependence is actually relevant, given the wide measurement intervals (2 and 5 years).

8. **Complex sample design considerations:**

   The NLTCS has a complex sampling design, including stratification, clustering and a complex enrollment (Clark, 1998). Part of the enrollment design problems has already being accounted for, by conditioning on the age of the subjects. There are, however, a number of characteristics that have been used to define sampling strata that can potentially be correlated with different ways of aging, like geographic location, that can be considered a proxy for socioeconomic condition and for exposure to different types of weather. Stratified sampling based on these characteristics can potentially bias the estimates of the population-level distribution of membership. A simple way of incorporating this information is as covariates, using the Grouped data model, from Chapter 4. Then, using external information about the distribution of those covariates in the population, we can integrate them out to get an unconditional model.

Bibliography

Airoldi, E. M., Blei, D. M., Fienberg, S. E., and Xing, E. P. (2008), "Mixed membership stochastic blockmodels," *The Journal of Machine Learning Research*, 9, 1981–2014.

Airoldi, E. M., Fienberg, S. E., Joutard, C., and Love, T. M. (2007), "Discovering Latent Patterns with Hierarchical Bayesian Mixed-Membership Models," *Data Mining Patterns: New Methods and Applications*, 240–275.

Bertolet, M. (2008), "To Weight Or Not To Weight? Incorporating Sampling Designs Into Model-Based Analyses," Ph.D. thesis, Carnegie Mellon University.

Blei, D. M. and Lafferty, J. D. (2007), "A correlated topic model of *Science*," *Annals of Applied Statistics*, 1, 17–35.

Blei, D. M., Ng, A., and Jordan, M. I. (2003), "Latent Dirichlet allocation," *Journal of Machine Learning Research*, 3, 993–1022.

Casella, G. and George, E. (1992), "Explaining the Gibbs sampler," *American Statistician*, 46, 167–174.

Chib, S. (2004), "Markov Chain Monte Carlo Technology," in *Handbook of Computational Statistics*, eds. Gentle, J., Härdle, W., and Mori, Y., Springer Berlin, chap. II.3, pp. 71 – 102.

Clark, R. (1998), “An Introduction to the National Long-Term Care Survey,” http://aspe.hhs.gov/daltcp/reports/nltcssu2.htm, office of Disability, Agigng, and Long-Term Care Policy with the U.S. Department of Health and Human Services.

Connor, J., Fienberg, S., Erosheva, E., and White, T. (2006), “Towards a restructuring of the national long term care survey: A longitudinal perspective,” Tech. rep.

Connor, J. T. (2006), “Multivariate Mixture Models to Describe Longitudinal Patterns of Frailty in American Seniors,” Ph.D. thesis, Department of Statistics & H. John Heinz III School of Public Policy & Management. Carnegie Mellon University.

Cooil, B. and Varki, S. (2003), “Using the conditional Grade-of-Membership model to assess judgment accuracy,” *Psychometrika*, 68, 453–471.

Corder, L. and Manton, K. (1991), “National surveys and the health and functioning of the elderly: the effects of design and content,” *Journal of the American Statistical Association*, 86, 513–525.

Cox, B. G. (2009), “Revised Cross Sectional Weights for the National Long-Term Care Survey.” in *The 2004 Wave of the National Long Term Care Survey: New Research Directions.*, The National Archive of Computerized Data on Aging, Bethesda, MD.

Devroye, L. (1986), *Non-Uniform Random Variate Generation*, New York: Springer-Verlag.

Dimatteo, I., Genovese, C., and Kass, R. (2001), “Bayesian curve-fitting with free-knot splines,” *Biometrika*, 88, 1055.

Erosheva, E. (2002), “Grade of membership and latent structures with application to disability survey data,” Ph.D. thesis, Department of Statistics. Carnegie Mellon University.

— (2005), “Comparing Latent Structures of the Grade of Membership, Rasch, and Latent

Class Models," *Psychometrika*, 70, 619–628.

Erosheva, E., Fienberg, S., and Joutard, C. (2007), "Describing disability through individual-level mixture models for multivariate binary data," *Annals of Applied Statistics*, 1, 502–537.

Erosheva, E., Fienberg, S., and Junker, B. (2002), "Alternative statistical models and representations for large sparse multi-dimensional contingency tables," 11, 485–505.

Erosheva, E., Fienberg, S. E., and Lafferty, J. D. (2004), "Mixed-membership models of scientific publications," *Proceedings of the National Academy of Sciences*, 101, 5220–5227.

Erosheva, E. and White, T. (2009), "Using 2004 NLTCS survey to explore issues in chronic disability measurement," in *The 2004 Wave of the National Long Term Care Survey: New Research Directions.*, The National Archive of Computerized Data on Aging, Bethesda, MD.

Erosheva, E. A. and White, T. A. (2010), "Issues in Survey Measurement of Chronic Disability: An Example from the National Long Term Care Survey," *Journal of Official Statistics*, 26, 317–339.

Ferguson, T. (1973), "A Bayesian analysis of some nonparametric problems," *The Annals of Statistics*, 1, 209–230.

Ferrucci, L., Guralnik, J., Simonsick, E., Salive, M., Corti, C., and Langlois, J. (1996), "Progressive versus catastrophic disability: a longitudinal view of the disablement process," *The Journals of Gerontology: Series A*, 51, M123.

Fienberg, S. and Mason, W. (1985), "Specification and implementation of age, period and cohort models," in *Cohort Analysis in Social Research*, eds. Fienberg and Mason, New York: Springer-Verlag.

Gamerman, D. (1997), *Markov Chain Monte Carlo*, Boca Raton, FL: Chapman & Hall New York.

Gelman, A., Carlin, J., Stern, H., and Rubin, D. (2004), *Bayesian data analysis.*, Boca Raton, FL: Chapman and Hall/CRC.

Geman, S. and Geman, D. (1984), “Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images,” *IEEE Trans. on PAMI*, 6, 721–741.

Gilks, W. and Best, N. (1992), “Adaptive rejection sampling for Gibbs sampling,” *Applied Statistics*, 41, 337–348.

Gilks, W. and Roberts, G. (1996), “Strategies for improving MCMC,” in *Markov chain Monte Carlo in practice*, eds. Gilks, W., Richardson, S., and Spiegelhalter, D., Chapman & Hall/CRC, pp. 89–114.

Goodman, L. A. (1974), “Exploratory latent structure analysis using both identifiable and unidentifiable models,” *Biometrika*, 61, 215–231.

Gormley, C. (2006), “Statistical Models For Rank Data,” Ph.D. thesis, Department of Statistics, University of Dublin, Trinity College.

Green, P. J. (1995), “Reversible jump Markov chain Monte Carlo computation and Bayesian model determination,” *Biometrika*, 82, 711–732.

Haberman, S. J. (1995), “Review: Statistical applications using fuzzy sets, by K. Manton, M. Woodbury and H. Tolley,” *Journal of the American Statistical Association*, 90, 1131–1133.

Holland, P. W. and Rosenbaum, P. R. (1986), “Conditional association and unidimensionality in monotone latent variable models,” *Annals of Statistics*, 14, 1523–1543.

Hougaard, P. (2000), *Analysis of Multivariate Survival Data*, New York: Springer Verlag.

Ibrahim, J., Chen, M., and Sinha, D. (2001), *Bayesian Survival Analysis*, New York:

Springer Verlag.

Kalbfleisch, J. and Prentice, R. (2002), *The Statistical Analysis of Failure Time Data*, New York: Wiley, 2nd ed.

Kreuter, F. and Muthén, B. (2008), "Analyzing criminal trajectory profiles: Bridging multilevel and group-based approaches using growth mixture modeling," *J Quant Criminol*, 24, 1–31.

Kurland, B. and Heagerty, P. (2005), "Directly parameterized regression conditioning on being alive: analysis of longitudinal data truncated by deaths," *Biostatistics*, 6, 241.

Kurland, B., Johnson, L., Egleston, B., and Diehr, P. (2009), "Longitudinal Data with Follow-up Truncated by Death: Match the Analysis Method to Research Aims," *Statistical science*, 24, 211.

Liu, J. (1999), "Markov chain Monte Carlo and related topics," Tech. rep., Department of Statistics. Standford University.

Manrique-Vallier, D. and Fienberg, S. (2008), "Population size estimation using individual level mixture models," *Biometrical Journal*, 50, 1051–1063.

Manton, K. G. (2008), "Recent Declines in Chronic Disability in the Elderly U.S. Population: Risk Factors and Future Dynamics," *Annual Review of Public Health*, 29, 91–113.

Manton, K. G., Corder, L., and Stallard, E. (1997), "Chronic disability trends in elderly United States populations: 1982-1994," *Proceedings of the National Academy of Sciences*, 94, 2593–2598.

Manton, K. G., Gu, X., Huang, H., and Kovtun, M. (2004), "Fuzzy set analyses of genetic determinants of health and disability status," *Statistical Methods in Medical Research*, 13, 395.

Manton, K. G., Gu, X., and Lamb, V. (2006), “Change in chronic disability from 1982 to 2004/2005 as measured by long-term changes in function and health in the US elderly population,” *Proceedings of the National Academy of Sciences*, 103, 18374.

Manton, K. G., Gu, X., and Lowrimore, G. R. (2008), “Cohort Changes in Active Life Expectancy in the U.S. Elderly Population: Experience From the 1982-2004 National Long-Term Care Survey,” *J Gerontol B Psychol Sci Soc Sci*, 63, S269–281.

Manton, K. G., Lamb, V., et al. (2007), “Medicare Cost Effects of Recent US Disability Trends in the Elderly: Future Implications,” *Journal of Aging and Health*, 19, 359.

Manton, K. G., Stallard, E., Liu, K., Manton, K., Singer, B., and Suzman, R. (1993), *Forecasting the Health of Elderly Populations*, New York: Springer-Verlag.

Manton, K. G., Stallard, E., and Woodbury, M. (1991), “A multivariate event history model based upon fuzzy states: Estimation from longitudinal surveys with informative nonresponse,” *Journal of Official Statistics*, 7, 261–293.

Manton, K. G., Woodbury, M., Stallard, E., and Corder, L. (1992), “The Use of Grade-of-Membership Techniques to Estimate Regression Relationships,” *Sociological Methodology*, 22, 321–381.

Manton, K. G., Woodbury, M. A., and Tolley, H. D. (1994), *Statistical Applications Using Fuzzy Sets*, New York: Wiley.

Marin, J.-M. and Robert, C. P. (2007), *Bayesian Core: A Practical Approach to Computational Bayesian Statistics (Springer Texts in Statistics)*, New York: Springer-Verlag.

Nagin, D. (1999), “Analyzing developmental trajectories: A semiparametric, group-based approach,” *Psychological Methods*, 4, 139–157.

Nagin, D. and Land, K. (1993), “Age, criminal careers, and population heterogeneity: Specification and estimation of a nonparametric, mixed Poisson model,” *Criminology*, 31, 327–362.

Robert, C. and Casella, G. (2004), *Monte Carlo Statistical Methods*, New York: Springer-Verlag, 2nd ed.

Robert, C. P. (2004), "Bayesian Computational Methods," in *Handbook of computational statistics*, eds. Gentle, J., Härdle, W., and Mori, Y., Springer Berlin, chap. III.11, pp. 719 – 7656.

Rubin, D. (1976), "Inference and missing data," *Biometrika*, 63, 581.

Ryder, N. B. (1985), "The cohort as a concept in the study of social change," in *Cohort Analysis in Social Research*, eds. Fienberg and Mason, New York: Springer-Verlag.

Sijtsma, K. and Junker, B. (2006), "Item response theory: Past performance, present developments, and future expectations," *Behaviormetrika*, 33, 75–102.

Stallard, E. (2005), "Trajectories of Morbidity, Disability, and Mortality among the US Elderly Population: Evidence from the 1984–1999 NLTCS," *North American Actuarial Journal*, 11.

Tanner, M. (1996), *Tools for Statistical Inference: Methods for the Exploration of Posterior Distributions and Likelihood Functions*, New York: Springer Verlag, 3rd ed.

Weibull, W. (1951), "A statistical distribution function of wide applicability," *Journal of applied mechanics*, 18, 293–297.

White, T. (2008), "Extensions of Latent Class Transition Models with Application to Chronic Disability Survey Data," Ph.D. thesis, University of Washington.

Woodbury, M., Clive, J., and Garson Jr, A. (1978), "Mathematical typology: A grade of membership technique for obtaining disease definition." *Computers in Biomedical Research*, 11, 277–98.