The Annals of Applied Statistics
2014, Vol. 8, No. 4, 2268–2291
DOI: 10.1214/14-AOAS769
© Institute of Mathematical Statistics, 2014

LONGITUDINAL MIXED MEMBERSHIP TRAJECTORY MODELS FOR DISABILITY SURVEY DATA[1]

BY DANIEL MANRIQUE-VALLIER

Indiana University

We develop methods for analyzing discrete multivariate longitudinal data and apply them to functional disability data on the U.S. elderly population from the National Long Term Care Survey (NLTCS), 1982–2004. Our models build on a Mixed Membership framework, in which individuals are allowed multiple membership on a set of extreme profiles characterized by time-dependent trajectories of progression into disability. We also develop an extension that allows us to incorporate birth-cohort effects, in order to assess inter-generational changes. Applying these methods, we find that most individuals follow trajectories that imply a late onset of disability, and that younger cohorts tend to develop disabilities at a later stage in life compared to their elders.

1. Introduction. This paper introduces new models and estimation procedures to analyze discrete multivariate longitudinal data on functional disability, motivated by the analysis of data from the National Long Term Care Survey (NLTCS). The NLTCS is a longitudinal panel survey instrument aimed at assessing chronic disability among the elderly (65+) population in the United States. It enables researchers to answer important questions related to the aging process and disability prevalence 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 the nature of disability changing for younger generations? [Connor et al. (2006)]. Answers to these questions are of 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, Lamb and Gu (2007)].

Many of the relevant public policy questions for which the NLTCS can potentially provide answers are related to 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 differently than people born 20 years before?"). Thus, to answer these questions we need to consider the longitudinal dimension of these data. In addition, as not everyone could be expected to age the exact same way, it is safe to assume that elderly

Received October 2012; revised April 2014.

[1]Supported in part by NIH Grant R01 AG023141-01 to Carnegie Mellon University and NSF Grant SES-11-31897 to Duke University.

Key words and phrases. NLTCS, Mixed Membership, trajectories, multivariate analysis, MCMC, cohort analysis.

American people constitute a heterogeneous population. Models for longitudinal disability data need to be capable of accounting for such heterogeneity.

Although the longitudinal nature of the NLTCS data is frequently invoked [e.g., Corder and Manton (1991), Manton, Corder and Stallard (1997), Manton, Gu and Lamb (2006)], efforts to analyze the data using true longitudinal methods have been few and far between. Most researchers have instead analyzed the NLTCS as a series of uncorrelated cross-sectional samples [see, e.g., Manton, Corder and Stallard (1997), Manton, Gu and Lamb (2006), Manton, Lamb and Gu (2007)]. Recent attempts to deal with the longitudinal nature of the NLTCS have been undertaken by Stallard (2005), Connor (2006) and White and Erosheva (2013).

The new models and methods that we propose in this paper, which we call *Trajectory Grade of Membership* models (TGoM), seek to capture both the longitudinal nature of the individual NLTCS data and the inherent individual heterogeneity of the aging process. These models handle individual heterogeneity using the concept of Mixed Membership [Erosheva and Fienberg (2005), Erosheva, Fienberg and Lafferty (2004)]. Mixed Membership models describe a small number of ideal types of individuals (or extreme profiles) and let each individual partially belong to each pure type, to a different degree. At the same time, TGoM models focus on the longitudinal nature of the process by defining the extreme profiles as typical progressions over time. We also introduce an extension to this model aimed at capturing differences across generational cohorts. We do this by allowing individuals' Mixed Membership to depend on their dates of birth.

The remainder of this article is organized as follows. In the next section we present a brief introduction and description of the National Long Term Survey. Next, in Section 3, we describe the basic TGoM model and its extension to handle generational cohorts. Estimation algorithms based on MCMC sampling are introduced in Section 4 and fully described in Appendices A and B. In Section 5 we apply the TGoM models to the NLTCS. Finally, in Section 6, we conclude with a discussion on the insights provided by the models, their limitations and possible extensions.

2. 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 the United States population aged 65 years or more [Corder and Manton (1991)]. It consists of six waves, conducted in 1982, 1984, 1989, 1994, 1999 and 2004. In very rough terms, each wave consists of interviews to approximately 20,000 people, from which around 15,000 are previously interviewed individuals. Each wave includes a fresh new sample of around 5000 individuals. These refreshment samples serve the double purpose of replacing those who have died since the previous wave and of keeping each wave representative of the current state of the population over 65 [Clark (1998)]. A total of around 49,000 people have been screened in the survey between 1982 and 2004.

The NLTCS assesses functional disability by evaluating subjects' ability to perform two sets of activities. The first one, called *Activities of Daily Living* (ADL), comprises basic self-care activities, such as bathing, eating and dressing. The second, *Instrumental Activities of Daily Living* (IADL), involves activities necessary for independent living within a community, like preparing meals or maintaining finances. The NTLCS determines the functional status in these activities through answers to a series of triggering questions, which are then summarized as binary response items that indicate the presence or absence of impairments.

The design of the NLTCS is such that the survey data can 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 first screens each sampled individual using a special, "screener," questionnaire aimed at quickly detecting if he or she is chronically disabled. The operational definition of "chronically disabled" in the context of the NLTCS requires that the individual presents an impairment in some ADL or IADL lasting or expected to last at least 90 days. If screened-out, the individual's status is registered and they are re-screened 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 individuals living in the community. 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 [Clark (1998)].

In what follows, we have used a subset of the NLTCS consisting of all six binary answers to questions about the individual's ability to perform ADLs (EAT: Eating; DRS: Dressing; TLT: Toileting; BED: Getting in and out of bed; MOB: Inside mobility; BTH: Bathing), from all six waves of the NLTCS. We obtained ages and dates of birth from linked Medicare data from the Centers for Medicare and Medicaid services (CMS). We provide further details about our data preprocessing in Section 5.

3. Mixed Membership trajectory models. The goal of this analysis is to characterize typical progressions in acquisition of disabilities over time while taking into consideration and characterizing the heterogeneity of the population. For this we combine two main ideas.

The first idea is clustering based on trajectories. This is the idea behind Latent Trajectory models [LTMs; Nagin (1999)]. Broadly speaking, LTMs are mixture models of the form

$$p(\mathbf{y}|\mathbf{x}) = \sum_{k=1}^{K} \pi_k f_k(\mathbf{y}|\mathbf{x}), \tag{3.1}$$

where $\mathbf{y}$ is a vector containing T longitudinal measurements of a response variable of interest and $\mathbf{x}$ is a vector that contains the corresponding T values of a

time-dependent covariate. The joint densities corresponding to each mixture component, $f_k(\cdot)$, are in turn modeled using parametric *trajectory functions*. Trajectory functions (or simply *trajectories*) describe typical progressions over time, usually modeling the dependence of the outcome variables as a function of age. For a given population, LTMs provide estimates of both the trajectories and the individuals' distribution over them. Therefore, LTMs perform data-driven clustering based on evolution over time [see Nagin (1999) for details]. Connor (2006) adapted this technique for the analysis of multivariate discrete data and applied it to the NLTCS. The trajectory curves represented the probability of presenting a disability as a function of age. This tool provides a simple and easy mechanism to interpret typical ways of aging, with a degree of heterogeneity handling. However, it assumes that individuals within a class are perfectly homogeneous. It thus attributes all the potential within-class variability to random fluctuations. In Connor's formulation, this assumption essentially says that, within a class, every single individual responds to the exact same underlying aging process. It thus disregards the fact that classes are ideal constructions to which possibly no actual individuals belong [Kreuter and Muthén (2008)].

The second idea, Mixed Membership, provides a powerful and conceptually attractive way of relaxing the within-class homogeneity assumption. Similarly to traditional clustering techniques, like the Latent Class model [Goodman (1974)] or LTMs, Mixed Membership models still assume the existence of a small number of classes, called *ideal types* or *extreme profiles*. However, instead of forcing every individual into one and only one class, they allow them to belong simultaneously to more than one, in different degree. The Grade of Membership model [GoM; Manton, Stallard and Woodbury (1991), Woodbury, Clive and Garson (1978)] is an example of a Mixed Membership model that has been successfully applied to the cross-sectional analysis of the NTLCS [see, e.g., Erosheva, Fienberg and Joutard (2007), Manton, Corder and Stallard (1997), Manton, Gu and Lamb (2006), Manton, Lamb and Gu (2007)]. Erosheva, Fienberg and Joutard (2007) developed a full Bayesian version of the GoM model and applied it to a pooled across-waves version of the NLTCS.

The approach we present here combines LTMs with Mixed Membership. It seeks to produce a soft clustering based on trajectories. Similar to LTM, it assumes that for a given population we can identify a few ways of progressing over time, which we consider ideal extreme cases. At the same time it assumes that individuals in the population do not exactly correspond to these typical profiles, but instead behave somewhere in between them, in quantifiable ways. Note that this approach is conceptually different from previous cross-sectional applications of the GoM model to the study of disability. In those applications extreme profiles represented ideal types of disability, whereas in TGoMs they represent ideal types of people. In the same way, it also differs from other previously proposed time-dependent Mixed Membership models, which specify time-evolving individual membership [e.g., Stallard (2005), Xing, Fu and Song (2010)]. In TGoMs the membership is an immutable characteristic of the individual.

3.1. *Basic TGoM model.* We consider a sample composed of N individuals. Following Mixed Membership ideas, we assume the existence of a number, K, of reference types of individuals called *extreme profiles*. These extreme profiles represent idealized individuals. This means that it might be the case that no real individual corresponds exactly to any of them. Instead, we assume that each individual $i = 1, \dots, N$ has an associated *membership vector*, $\mathbf{g}_i = (g_{i1}, \dots, g_{iK})$, whose kth component, g_{ik}, represents their *degree of membership* into the kth extreme profile. We constrain membership vectors so that their components are positive numbers that sum to 1, that is, they lie on a $(K-1)$-dimensional unit simplex, Δ_{K-1}. In this way, we identify ideal individuals of the kth type as those whose membership vectors' components are zeros on each component distinct from k, and $g_{ik} = 1$. For instance, we say that an individual with membership vector $\mathbf{g}_i = (0, 0, 1, 0)$ belongs exclusively to the extreme profile $k = 3$. Similarly, we can represent more complex membership structures. For example, $\mathbf{g}_i = (0.1, 0.2, 0.4, 0.3)$ indicates that individual i has 10% membership in the first extreme profile, 20% in the second and so on.

We are interested in modeling the progression of disability as time passes. We start by modeling ideal individuals. Let individual i provisionally be a full member of extreme profile k, that is, $g_{ik} = 1$. Let $y_{ij}(\tau)$ be 1 if the individual does experience difficulties performing ADL j at age τ, and 0 otherwise. We model the evolution of the probability of a positive response to question j, $y_{ij}(\tau)$, as a function of age, $\lambda_{jk}(\tau)$, so that

$$\lambda_{jk}(\tau) = \Pr\big(y_{ij}(\tau) = 1 | g_{ik} = 1, \boldsymbol{\beta}_{jk}, \tau\big). \tag{3.2}$$

Here $\boldsymbol{\beta}_{jk}$ is a generic vector of parameters that indexes $\lambda_{jk}(\cdot)$ within a parametric family, for example, the parameters of a linear logistic curve. We call the functions $\lambda_{jk}(\cdot)$ *extreme trajectories*.

Now moving to actual individuals, we specify the corresponding trajectory of a generic, nonideal individual i, with membership vector $\mathbf{g}_i = (g_{i1}, \dots, g_{iK})$, as the convex combination

$$\lambda_j^{(i)}(\tau) = \Pr\big(y_{ij}(\tau) = 1 | \mathbf{g}_i, \boldsymbol{\beta}_j, \tau\big) = \sum_{k=1}^{K} g_{ik} \lambda_{jk}(\tau),$$

where $\boldsymbol{\beta}_j = (\boldsymbol{\beta}_{j1}, \dots, \boldsymbol{\beta}_{jK})$.

Although τ is a continuously-varying quantity, we only have measurements at each of the $t = 1, \dots, T = 6$ occasions, corresponding to the waves of the survey. Thus, we define $y_{ijt} = y_{ij}(\mathrm{Age}_{it})$, where Age_{it} is the age of individual i at measurement time $t = 1, \dots, T$. We group these numbers into individual vectors $\mathbf{Age}_i = (\mathrm{Age}_{i1}, \dots, \mathrm{Age}_{iT})$. Then we have that

$$\begin{aligned} p(y_{ijt} | \mathbf{g}_i, \boldsymbol{\beta}_j, \mathbf{Age}_i) &= \mathrm{Bern}\big(y_{ijt} | \lambda_j^{(i)}(\mathrm{Age}_{it})\big) \\ &= \sum_{k=1}^{K} g_{ik} \, \mathrm{Bern}\big(y_{ijt} | \lambda_{jk}(\mathrm{Age}_{it})\big), \end{aligned}$$

where $\mathrm{Bern}(y|p) = p^y(1-p)^{1-y}$, for $y \in \{0, 1\}$ and $0 < p < 1$.

Next, we assume that, for a single individual, the J responses at each of the T measurement times are conditionally independent of one another, given their membership vector, $\mathbf{g}_i$, and covariate vector $\mathbf{Age}_i$. Under this assumption we effectively use the membership vector and the covariates to decouple the dependence structure present in the components of the response. Then we have

$$p(\mathbf{Y}_i|\mathbf{g}_i, \boldsymbol{\beta}, \mathbf{Age}_i) = \prod_{j=1}^{J}\prod_{t=1}^{T}\sum_{k=1}^{K} g_{ik}\,\mathrm{Bern}\big(y_{ijt}|\lambda_{jk}(\mathrm{Age}_{it})\big), \tag{3.3}$$

where $\mathbf{Y}_i = (y_{ijt})_{J=1,\dots,J,t=1,\dots,T}$ and $\boldsymbol{\beta} = (\boldsymbol{\beta}_1, \dots, \boldsymbol{\beta}_J)$. By assuming that each individual has been randomly sampled from the population, we finally get the joint model of $\mathbf{Y} = (\mathbf{Y}_i)$, conditional on $\mathbf{g} = (\mathbf{g}_i)$ and $\mathbf{Age} = (\mathbf{Age}_i)$,

$$p(\mathbf{Y}|\mathbf{g}, \boldsymbol{\beta}, \mathbf{Age}) = \prod_{i=1}^{N}\prod_{j=1}^{J}\prod_{t=1}^{T}\sum_{k=1}^{K} g_{ik}\,\mathrm{Bern}\big(y_{ijt}|\lambda_{jk}(\mathrm{Age}_{it})\big). \tag{3.4}$$

We assume that membership vectors are i.i.d. samples from a common distribution $G_{\boldsymbol{\alpha}}$, with support on the simplex Δ_{K-1}. This yields the unconditional (on $\mathbf{g}$) model for the sample $\mathbf{Y}$,

$$p(\mathbf{Y}|\boldsymbol{\beta}, \mathbf{Age}) = \prod_{i=1}^{N}\int_{\Delta_{K-1}}\prod_{j=1}^{J}\prod_{t=1}^{T}\sum_{k=1}^{K} \omega_k\,\mathrm{Bern}\big(y_{ijt}|\lambda_{jk}(\mathrm{Age}_i)\big)G_{\boldsymbol{\alpha}}(d\boldsymbol{\omega}), \tag{3.5}$$

where $\boldsymbol{\omega} = (\omega_1, \dots, \omega_K) \in \Delta_{K-1}$. Figure 1 shows a graphical representation of the structure of this model.

As Erosheva, Fienberg and Joutard (2007) discuss for the Grade of Membership model, the model in (3.3) admits the augmented data representation,

$$\begin{aligned} &f^{\mathrm{AUG}}(\mathbf{Y}_i, \mathbf{Z}_i|\mathbf{Age}_i, \boldsymbol{\beta}, \mathbf{g}_i) \\ &\quad = \prod_{j=1}^{J}\prod_{t=1}^{T}\prod_{k=1}^{K}\big[g_{ik}\,\mathrm{Bern}\big(y_{ijt}|\lambda_{jk}(\mathrm{Age}_{it})\big)\big]^{I(z_{ijt}=k)}, \end{aligned} \tag{3.6}$$

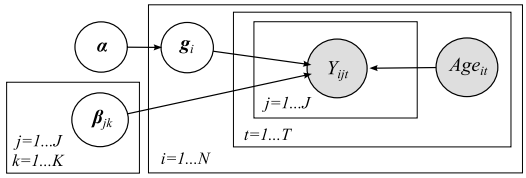


FIG. 1. *Graphical probabilistic representation of the basic TGoM model. Observed variable* Age_{it} *is the age of individual* i *at survey wave* t. *Gray nodes represent observed quantities*; *white nodes represent parameters to estimate.*

where $\mathbf{Z}_i = (z_{ijt})_{j=1,\dots,J,t=1,\dots,T}$ with $z_{ijt} \in \{1, 2, \dots, K\}$. Following Erosheva, Fienberg and Joutard (2007), it is easy to show that the expression in (3.3) is equivalent to

$$p(\mathbf{Y}_i|\mathbf{Age}_i, \boldsymbol{\beta}, \mathbf{g}_i) = \sum_{\mathbf{z}\in\mathcal{Z}} f^{\mathrm{AUG}}(\mathbf{Y}_i, \mathbf{z}|\mathbf{Age}_i, \boldsymbol{\beta}, \mathbf{g}_i), \tag{3.7}$$

where $\mathcal{Z} = \{(z_{jt})_{J\times T} : z_{jt} \in \{1, \dots, K\}\}$. We then see that the model in (3.3) can be thought of as a marginalized version of the model in (3.6). This equivalence shows that the TGoM model conforms to the general mixed-membership structure described in Erosheva, Fienberg and Lafferty (2004). It also makes it possible to construct algorithms for posterior inference of the TGoM using the augmented model [Tanner (1996)].

3.2. *Detailed specification*. The extreme trajectories functions, $\lambda_{jk}(\cdot)$, encode several assumptions about the dynamics of the underlying process over time. Thus, their specific functional form must be application-specific. For this application to the NLTCS, following Connor (2006), we use a linear logit specification

$$\mathrm{logit}\big[\lambda_{jk}(\tau)\big] = \beta_{0jk} + \beta_{1jk}\tau. \tag{3.8}$$

Here $\boldsymbol{\beta}_{jk} = (\beta_{0jk}, \beta_{1jk})$. This specification expresses the intuitively sound notion that the underlying probability of disability is a monotonic function of age. It also has the advantage of being relatively simple, with just $2 \times J$ parameters per extreme profile. In the supplementary material [Manrique-Vallier (2014)] we present an analysis using an alternative specification and include a discussion about the appropriateness of 3.8.

Similar to Erosheva (2002) and Airoldi et al. (2008), we take the common distribution of the N membership vectors $\mathbf{g}_i$, $G_{\boldsymbol{\alpha}}$, as

$$\mathbf{g}_i|\boldsymbol{\alpha} \overset{\text{i.i.d.}}{\sim} \mathrm{Dirichlet}(\boldsymbol{\alpha}), \tag{3.9}$$

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

The Dirichlet distribution has some good properties in this setting. First, it is conjugate to the multinomial distribution. This simplifies computations using Gibbs samplers. Second, adopting the reparametrization $\boldsymbol{\alpha} = (\alpha_0 \cdot \xi_1, \dots, \alpha_0 \cdot \xi_K)$ with $\alpha_0 > 0$, $\xi_k > 0$ and $\sum_k \xi_k = 1$, we can interpret the vector $\boldsymbol{\xi} = (\xi_1, \dots, \xi_K)$ as the average proportion of responses generated by the kth extreme profile and α_0 as a parameter governing the spread of the distribution: as α_0 approaches 0, the samples from $G_{\boldsymbol{\alpha}}$ are more and more concentrated on the vertices of the simplex Δ_{K-1}; and as α_0 increases they are more concentrated near its mean, $\boldsymbol{\xi}$.

As Erosheva, Fienberg and Joutard (2007) and Airoldi et al. (2007) discuss, a priori setting parameter $\boldsymbol{\alpha}$ in the Dirichlet distribution is too strong an assumption to do realistic modeling. Estimates can be highly sensitive to this prior specification. For this reason, we prefer to estimate these parameters directly from the

data, specifying hyperpriors and computing posterior distributions. We specify hyperpriors for α_0 and $\boldsymbol{\xi}$ similar to Erosheva (2002) and Erosheva, Fienberg and Joutard (2007): $\alpha_0 \sim \text{Gamma}(a_\alpha, b_\alpha)$ and $\boldsymbol{\xi} \sim \text{Dirichlet}(\mathbf{1}_K)$. This specification takes advantage of the interpretation of the parameters α_0 and ξ, considering them as independent entities and modeling them separately. For the same reason we also assume that $p(\alpha_0, \boldsymbol{\xi}) = p(\alpha_0)p(\boldsymbol{\xi})$.

We specify the priors for the parameters that define the extreme trajectories, $\boldsymbol{\beta}_{jk} = (\beta_{0jk}, \beta_{jk})$, as two independent normal distributions, $\beta_{0jk} \overset{\text{i.i.d.}}{\sim} N(\mu_0, \sigma_0^2)$ and $\beta_{1jk} \overset{\text{i.i.d.}}{\sim} N(\mu_1, \sigma_1^2)$, for all $j = 1, \dots, J$ and $k = 1, \dots, K$. These priors can be set to be noninformative, by a priori specifying high variances. We also assume that $\boldsymbol{\beta}_{jk}$ are a priori independent of $\boldsymbol{\alpha}$.

3.3. *Representing generational changes.* The basic TGoM model from Section 3.1 takes advantage of the longitudinal nature of the NLTCS by following individuals as they age. It, however, attributes all variation over time, including changes in prevalence of disability patterns, to the individual progression of aging. Thus, it attributes all changes in prevalence of disability between different epochs to the aggregation of individuals that are at distinct points of their life trajectories.

To answer questions about changes in the ways of aging across different generations—for example, "are younger generations acquiring disabilities differently than older ones?"—we need to take into account the birth cohort of individuals. We do so by modeling the dependence between cohorts and the membership scores, keeping the extreme trajectories the same for the whole population. This arrangement allows us to read differences in the ways of aging as differences in the underlying distribution of membership, conditional on birth cohort. We interpret these differences using the common frame of reference provided by the extreme trajectories.

A direct way of enabling inter-generational comparisons under this framework is to keep the individual-level structure proposed for the basic TGoM model, but replace the common distribution of membership vectors with a family indexed by a function of the date of birth (DOB) covariate:

$$p(y_{ijt}|\mathbf{g}_i, \text{Age}_{it}, \boldsymbol{\beta}) = \sum_{k=1}^{K} g_{ik}\, \text{Bern}\big(y_{ijt}|\lambda_{jk}(\text{Age}_{it})\big),$$

$$\mathbf{g}_i|\text{DOB}_i \overset{\text{indep}}{\sim} G_{\boldsymbol{\alpha}(\text{DOB}_i)}.$$

For our application we keep the Dirichlet specification, but replace its parameter $\boldsymbol{\alpha}$ with a function of DOB, so that $G_{\boldsymbol{\alpha}(\text{DOB})} = \text{Dirichlet}(\boldsymbol{\alpha}_{(\text{DOB})})$. We note that under this specification the membership vectors, $\mathbf{g}_i$, are now dependent on a covariate.

A simple, yet reasonably flexible, way of specifying $\boldsymbol{\alpha}(\text{DOB})$ is by defining a number of cohorts and making it constant within each of them. Let $\Gamma =$

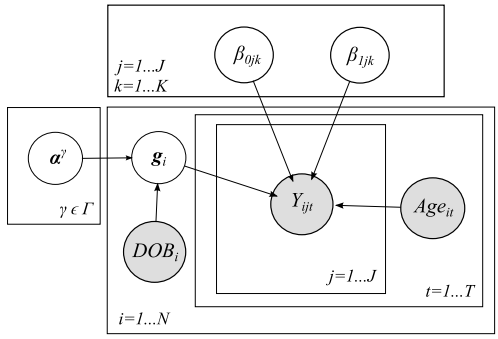


FIG. 2. *Probabilistic graphical representation of the extended TGoM model with cohort effects. Gray nodes represent observed quantities; white nodes represent parameters to estimate.*

$\{\gamma_1, \gamma_2, \dots, \gamma_C\}$ be a finite partition (contiguous nonoverlapping intervals) of the range of possible dates of birth. Define $\boldsymbol{\alpha}(\text{DOB}) = (\alpha_1(\text{DOB}), \alpha_2(\text{DOB}), \dots, \alpha_K(\text{DOB}))$ by

$$\alpha_k(\text{DOB}) = \prod_{\gamma \in \Gamma} (\alpha_k^{\gamma})^{I(\text{DOB} \in \gamma)}, \tag{3.10}$$

where $\alpha_k^{\gamma} > 0$. Then, we extend the TGoM model to handle cohort information by replacing the population level distribution of membership vectors, $p(\mathbf{g}_i|\boldsymbol{\alpha})$, with its conditional version, $p(\mathbf{g}_i|\boldsymbol{\alpha}(\text{DOB}_i))$. Figure 2 shows a graphical representation of this expanded model.

We specify the same hyperprior distribution that we used for the basic TGoM for all the newly introduced parameters. To this end, define $\alpha_0^{\gamma} = \sum_{k=1}^{K} \alpha_k^{\gamma}$ and $\xi_k^{\gamma} = \alpha_k^{\gamma}/\alpha_0^{\gamma}$, and take $\alpha_0^{\gamma} \overset{\text{i.i.d.}}{\sim} \text{Gamma}(\tau, \eta)$ and $\boldsymbol{\xi}^{\gamma} = (\xi_1^{\gamma}, \dots, \xi_K^{\gamma}) \overset{\text{i.i.d.}}{\sim} \text{Dirichlet}(\mathbf{1}_K)$.

4. Estimation. We developed MCMC algorithms based on Gibbs sampling to obtain samples from the posterior distribution of parameters for both the basic and the generational model. These algorithms rely on the augmented data representation in (3.6). We present the full description in Appendices A and B.

5. Application to the NLTCS. We have selected an extract from the NLTCS data that includes data from all six waves. These data include all the individuals that received the screener in at least one of the first five waves of the survey (1982, 1984, 1989, 1994 or 1999). We excluded individuals who entered the sample for the first time in 2004 because of lack of information about their dates of birth and death. Similarly, we excluded all the individuals that were institutionalized in 1982 because the NLTCS did not register their ADL statuses that year. The resulting sample size was $N = 38{,}428$ subjects. For each individual at each wave we focused on six ADLs: Eating ($j = 1$), Dressing ($j = 2$), Toileting ($j = 3$), Getting In or Out

TABLE 1
Cohort definition and distribution by wave

		Wave					
Cohort	**DOB**	**1982 ($t = 1$)**	**1984 ($t = 2$)**	**1989 ($t = 3$)**	**1994 ($t = 4$)**	**1999 ($t = 5$)**	**2004* ($t = 6$)**
1	–1906	6329	6025	1347	1397	617	70
2	1906–1914	7631	7082	3452	3335	1753	575
3	1914–1919	3696	7839	2627	5102	3679	2010
4	1919–1926	1	463	2410	4581	4724	3505
5	1926–	0	0	0	2478	6403	4251

*Only individuals present in 1999.

of Bed ($j = 4$), Inside Mobility ($j = 5$) and Bathing ($j = 6$). We determined the age of each individual in years by computing the difference between the interview and birth dates, and assuming 365 days for all years. For computing and prior specification purposes, we recentered ages at 80 years. However, for clarity we report any estimates or descriptive statistics related to age without the offset.

We defined five cohorts or generational groups, partitioning the ranges of possible dates of birth according to the intervals defined in the second column of Table 1. We selected these intervals so that they group approximately the same number of individuals. 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. Also note that neither the oldest (cohort 1—born before 1906) nor the youngest (cohort 5—born after 1926) cohorts span the whole range of relevant dates of birth in the NLTCS. In fact, the oldest individual in cohort 5 could be at most 78 years old in 2004, while the youngest individual from cohort 1 could not be younger than 76 years old in 1982.

5.1. *Basic GoM trajectory model.* We fitted the basic model described in Section 3.2 to the NLTCS data using the MCMC algorithm from Appendix A, for $K = 2, 3, 4$ and 5 extreme profiles.

We set the prior distribution for the proportions parameter of the membership vector, $\boldsymbol{\xi}$, as a uniform distribution over Δ_{K-1}, or Dirichlet($\mathbf{1}_K$). We specified the prior distribution for the corresponding concentration parameter, α_0, as Gamma$(1, 5)$, in shape/inverse scale parametrization. This last specification expresses a slight preference for small values of α_0, although not very pronounced. This choice is more a modeling decision than an expression of prior knowledge: small values of α_0 in the Dirichlet parametrization have the effect of concentrating the probability of individual membership vectors around the vertices of the unit simplex. This has the effect of producing individual membership vectors where

one single profile is predominant, but where the other profiles still exert some effect. This behavior is a desirable characteristic from an interpretative standpoint that allows us to discuss "predominant" profiles, while still having a significant degree of flexibility in the handling of heterogeneity due to the influence of the other profiles. For the parameters governing the extreme trajectories, we selected diffuse independent normal priors with $\mu = 0$ and variance $\sigma^2 = 100$.

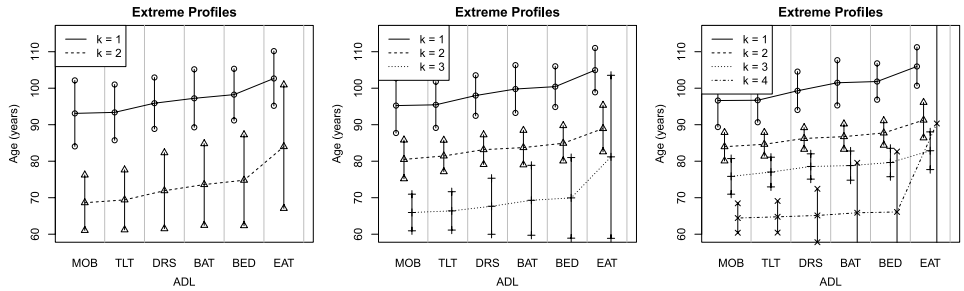
In all cases, the MCMC chains converged rapidly, reaching stationary distributions after approximately 15,000 iterations. Still, run times were long due to the chains' slow mixing. In all cases, we ran 120,000 iterations, discarded the first 20,000 and subsampled them, keeping 20% of the remaining. Similar to other latent variable models, the TGoM is invariant to permutation of its extreme profile labels. Thus, we inspected the trace plots for signs of label switching [Jasra, Holmes and Stephens (2005)]. No switching was found. Although label switching is a potential problem, in this application the modal regions of the posterior distributions seem to be well separated due to the abundance of data.

5.1.1. *Basic model results*. The basic TGoM model includes parameters that represent two distinct structural features: typical ways of aging, given by the extreme profiles (parameters $\boldsymbol{\beta}$), and the way individuals distribute with respect to these extreme profiles (parameters $\boldsymbol{\xi}$ and α_0). Extreme profile parameters can be difficult to interpret directly. Thus, we instead consider the quantities given by the transformation

$$\text{Age}_{q,jk} = -\frac{1}{\beta_{1jk}}\left[\beta_{0jk} + \log\left(\frac{1-q}{q}\right)\right] + 80, \tag{5.1}$$

for $q = 0.1$, $q = 0.5$ and $q = 0.9$. These parameters express the age at which an ideal individual of the extreme profile k reaches a probability q of being unable to perform ADL j. The 80 year offset is required because we have recentered the age data, subtracting 80. We also relabel extreme profiles according to the decreasing sequence of posterior estimates of ξ_k. This is necessary because of the TGoM's invariance to permutations of the extreme profile labels. This way, the expression "first extreme profile" ($k = 1$) will always refer to the extreme profile with the highest relative importance in the population (the one to which most individuals are the closest; see Section 5.1) and "the last" ($k = K$) to the one with the lowest.

Figure 3 and Table 2 present summaries of the posterior distribution of the extreme profile and Mixed Membership parameters, respectively. Plots in Figure 3 are based on posterior means of the quantities $\text{Age}_{q,jk}$, for models with $K = 2, 3, 4$. For each extreme profile, vertical line segments represent the age interval at which the probability of being unable to perform each ADL increases from 10% to 90%, that is, $[\text{Age}_{0.1,jk}, \text{Age}_{0.9,jk}]$. To aid visualization, we sorted the ADLs according to the $\text{Age}_{0.5,jk}$ estimates. Note that this procedure resulted in the exact same sequence of ADL in every case. Table 2 shows the posterior summaries of the Mixed Membership distribution parameters, α_0 and $\boldsymbol{\xi}$.



(a) $K = 2$ (b) $K = 3$ (c) $K = 4$

FIG. 3. *Posterior estimates of extreme profiles for models with* $K = 2, 3, 4$. *Vertical segments represent the age range at which ideal individuals' probabilities of disability go up from* 0.1 *to* 0.9, *for each ADL* ($[\text{Age}_{0.1,jk}, \text{Age}_{0.9,jk}]$). *For visualization purposes, ADLs are sorted according to* $\text{Age}_{0.5,jk}$ *posterior estimates.*

Estimates of the parameter α_0, in Table 2, are relatively small for all models. This was expected since the prior distribution of α_0, Gamma$(1, 5)$, was already expressing strong a priori preference for small values of α_0. However, as we can note from their very small posterior dispersion relative to the prior dispersion, these estimates are strongly data driven. This is not surprising, considering the large amount of data available to perform the estimations.

For all models, the extreme profile with the highest relative importance in the population, $k = 1$, represents a pattern of healthy aging, with a very late onset of disability. Extreme trajectories in this profile show that for any ADL, ideal individuals in this class have a very small probability of experiencing disability until approximately age 90. The remaining extreme profiles show patterns with progressively earlier onsets of disability, as we consider the extreme profiles in sequence. This is a feature worth noting: all models point to an inverse relationship between the relative importance of a profile in the population and its implied age of onset

TABLE 2

Posterior estimates of population-level parameters for basic model with $K = 2, 3, 4, 5$ *extreme profiles. Numbers between parenthesis are posterior standard deviations*

	α_0	ξ_1	ξ_2	ξ_3	ξ_4	ξ_5
$K = 2$	0.328	0.824	0.176	–	–	–
	(0.007)	(0.002)	(0.002)			
$K = 3$	0.261	0.645	0.251	0.104	–	–
	(0.006)	(0.004)	(0.004)	(0.002)		
$K = 4$	0.237	0.540	0.259	0.124	0.078	–
	(0.006)	(0.005)	(0.004)	(0.003)	(0.002)	
$K = 5$	0.235	0.496	0.244	0.128	0.074	0.058
	(0.005)	(0.007)	(0.006)	(0.003)	(0.002)	(0.001)

of disability. That is, most people's aging trajectories are closer to a profile that describes a late onset of disability.

We note that the sequence of ADLs obtained from sorting them according to their implied age of onset of disability (represented by parameter $\text{Age}_{0.5,jk}$) is the same for all extreme profiles of all models. Closer inspection reveals that the pattern of acquisition of disabilities directly inferred from the data closely follows what we can interpret as a sequence of activities decreasingly sorted in terms of difficulty: inside mobility, toileting, dressing, bathing, getting in and out of bed, and eating.

Another salient feature of these results is that for $k = 1, 2, 3$ and 4, the inferred slope parameters of the extreme trajectories (β_{1jk}) are all positive, even though the prior specification allows for negative values. This result supports the intuition that the probability of experiencing a disability in any ADL can only increase as one ages. It also makes it possible to construct the graphical summaries in Figure 3. The only exception to this regularity is in profile $k = 5$, in the model with $K = 5$ extreme profiles. In this profile trajectories exhibit a counterintuitive decreasing progression. We note that the relative importance of this profile in the population is small, with $\hat{\xi}_5 \approx 0.058$ (compared with $\hat{\xi}_1 \approx 0.496$ for the most important profile). From a modeling perspective, an obvious way of avoiding this type of aberrant behavior is to make it an impossibility a priori, restricting the support of the slope parameters to positive values. We have implemented such a model. However, while the rest of the parameters remained almost the same, the slope parameter of most trajectories in this profile were zero or very close to zero. These outcomes—together with the results obtained using a different trajectory specification, in the supplementary material [Manrique-Vallier (2014)]—suggest that this profile captures a small residual variability, which is not correctly modeled by the main extreme trajectories. Accounting for this effect is an area for future improvements.

To better understand the way TGoM models handle individual-level heterogeneity, it is instructive to visualize, in addition to the extreme trajectories, the actual individual trajectories that result from the individual-level mixing, $\lambda_{jk}^{(i)}(\tau)$. Plots in Figures 4 and 5 show a random sample of 100 such curves, overlaid over the three extreme trajectory curves, for each ADL, under the model with $K = 3$ and $K = 4$ extreme profiles, respectively. We see that most of the individual curves cluster in the vicinity of the extreme curves. This is expected, given the small value of the concentration parameter, α_0. However, we also see that a significant portion of the individual curves lies somewhere in between extremes, exhibiting trajectories that are the product of the interaction of more than one extreme. In particular, we observe a fair number of individual trajectories that fall in between extremes $k = 1$ and $k = 2$. These trajectories form a somewhat homogeneous cluster different from the extreme profiles. Nonetheless, the TGoM model has been able to accommodate them as a combination of (mainly) profiles $k = 1$ and $k = 2$, without needing to

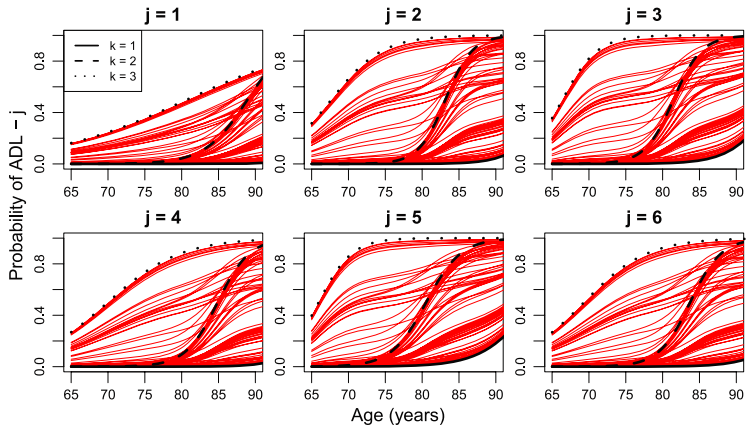


FIG. 4. *Individual-level mixture of trajectories for model with $K = 3$ extreme profiles for each ADL. Extreme trajectories are represented with thick lines and a random sample of* 100 *individual posterior trajectory curves are plotted using thin lines.*

create a whole new category for them. This behavior is what gives TGoM models the flexibility to accommodate complex individual heterogeneity while at the same time producing meaningful and interpretable summaries. Different from traditional LTMs [Connor (2006), Nagin (1999)], which require that individuals follow one and only one of the typical trajectories, this approach allows them to depart from the main tendencies, but not too much, thus retaining interpretability.

5.1.2. *Multivariate model diagnostics.* TGoM models explicitly model individual-level dependency between disability outcomes, both longitudinally and be-

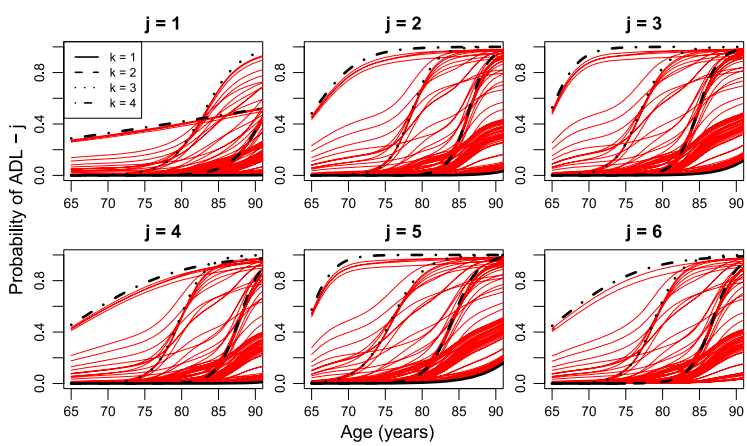


FIG. 5. *Individual-level mixture of trajectories for model with $K = 4$ extreme profiles for each ADL. Extreme trajectories are represented with thick lines and a random sample of* 100 *individual posterior trajectory curves are plotted using thin lines.*

tween ADLs, with the help of an individual-level Mixed Membership structure. In order to investigate empirically how TGoMs handle this dependency, we evaluate posterior univariate and multivariate out-of-sample predictive quantities. We define

$$\phi_{ijt}^K = \Pr(y_{ijt}^* = y_{ijt} | \mathcal{D}, K), \tag{5.2}$$

$$\phi_{ij}^K = \Pr(y_{ijt}^* = y_{ijt}, \text{ for all } t | \mathcal{D}, K), \tag{5.3}$$

$$\phi_{it}^K = \Pr(y_{ijt}^* = y_{ijt}, \text{ for all } j | \mathcal{D}, K), \tag{5.4}$$

$$\phi_{i}^K = \Pr(y_{ijt}^* = y_{ijt}, \text{for all } t \text{ and all } j | \mathcal{D}, K), \tag{5.5}$$

where y_{ijt}^* is the posterior predictive outcome of individual i in ADL-j at wave t, $\mathcal{D}$ are the NLTCS data, and K refers to the number of extreme profiles. Thus, for individual i, ϕ_{ijt}^K is the (univariate) posterior probability of correctly predicting outcome y_{ijt} using a TGoM with K extreme profiles; ϕ_{ij}^K is the probability of *simultaneously* correctly predicting the whole sequence of responses to ADL-j, at all waves; ϕ_{jt}^K is the corresponding probability of correctly predicting all the ADLs at wave t; and ϕ_i is the probability of simultaneously correctly predicting all the responses of an individual. In order to estimate the out-of-sample predictive performance of our models, we compute all these quantities using a 4-fold cross-validation scheme [Airoldi et al. (2010), Hastie, Tibshirani and Friedman (2009)].

As a comparison we also fit a model that assumes stochastic independence between univariate outcomes, given age. We fit six (one for each ADL) nonparametric logistic regressions of $\pi_{ijt} = \Pr(y_{ijt} = 1)$ on age, using Generalized Additive Models [GAM; Hastie, Tibshirani and Friedman (2009), Chapter 9]. We use this model as a reference for assessing how our models handle the multivariate structure present in the data. To this end, we compute quantities analogous to (5.2)–(5.5): $\phi_{ijt}^{\text{GAM}} = \text{Bern}(y_{ijt} | \hat{\pi}_{ijt})$, $\phi_{ij}^{\text{GAM}} = \prod_t \phi_{ijt}^{\text{GAM}}$, $\phi_{it}^{\text{GAM}} = \prod_j \phi_{ijt}^{\text{GAM}}$, and $\phi_i^{\text{GAM}} = \prod_{j,t} \phi_{ijt}^{\text{GAM}}$, where $\hat{\pi}_{ijt}$ is the fitted value of π_{ijt}. We compute these quantities using the same 4-fold cross-validation scheme we use for the TGoM quantities.

Table 3 shows the 4-fold cross-validated means (over all their subindexes) of ϕ_{ijt}, ϕ_{ij}, ϕ_{it} and ϕ_i, for TGoM with $K = 1, 2, \dots, 5$ extreme profiles and for the logistic GAM models. We take these numbers as estimates of the corresponding file-level rates of correct predictions for each model. We note that both TGoM and GAM models have similar univariate prediction rates, of around 80%, slightly favoring GAM. However, most of the joint prediction rates of the TGoM models with $K > 1$ are substantially better than the alternative. In particular, TGoM correct prediction rates for complete individual outcomes vectors, $\overline{\phi_i}$, are between 41.4% and 45.4%, while for the GAM model it drops down to 24.7%. We also observe that multivariate prediction rates using TGoM models tend to be much closer to their univariate prediction rates than the corresponding quantities using

TABLE 3

Out-of-sample rates of univariate and multivariate correct predictions for TGoM and nonparametric logistic regression models. Percentages between parentheses are the ratio of each entry with respect to its corresponding univariate correct prediction rates, $\overline{\phi_{ijt}}$

Model	$\overline{\boldsymbol{\phi_{ijt}}}$	$\overline{\boldsymbol{\phi_{ij}}}$	$\overline{\boldsymbol{\phi_{it}}}$	$\overline{\boldsymbol{\phi_{i}}}$
TGoM $K=1$	0.811 (100%)	0.644 (79.4%)	0.452 (55.7%)	0.251 (30.9%)
TGoM $K=2$	0.803 (100%)	0.666 (82.9%)	0.567 (70.6%)	0.414 (51.6%)
TGoM $K=3$	0.802 (100%)	0.668 (83.3%)	0.593 (73.9%)	0.440 (54.9%)
TGoM $K=4$	0.801 (100%)	0.668 (83.4%)	0.605 (75.5%)	0.451 (56.3%)
TGoM $K=5$	0.799 (100%)	0.664 (83.1%)	0.607 (76.0%)	0.454 (56.8%)
GAM-logistic	0.812 (100%)	0.645 (79.4%)	0.451 (55.5%)	0.247 (30.4%)

the GAM alternative. For instance, the ratio $\overline{\phi_i}/\overline{\phi_{ijt}}$ (numbers between parenthesis in the 5th column of Table 3) ranges from 51.6% to 56.8% for TGoM models ($K=2$ and $K=5$, resp.), while for GAM it falls down to 30.4%. We finally observe that estimates with the TGoM model with $K=1$ are almost identical to those obtained with GAM. This is because fitting TGoMs with only one extreme profile ($K=1$) is equivalent to fitting J independent logistic regressions of the response variable (ADLs) on the predictors (Age).

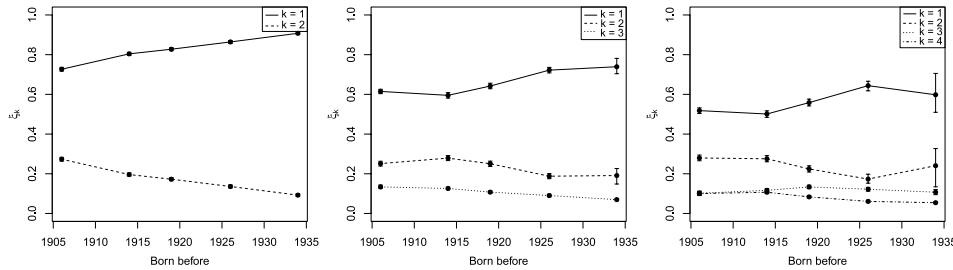
An interesting feature in Table 3 is that longitudinal predictions ($\overline{\phi_{ij}}$) are better than cross-sectional predictions ($\overline{\phi_{it}}$) for all models. This can be explained noting that both the TGoM and GAM approaches exploit the extra longitudinal information provided by the vectors of individuals' ages. By contrast, when modeling the multivariate cross-sectional structure, TGoM relies only on Mixed Membership and GAM only on independence given age. Nonetheless, the comparison between the two modeling approaches still favors TGoM models.

The conclusion of this prediction exercise is that TGoM models with more than one extreme profile do capture a large portion of the multivariate structure present in the data, both longitudinally and cross-sectionally.

5.2. *Fitting the cohort extensions.* We have fitted the model with extensions to handle cohort information to the NLTCS data using the MCMC algorithms in Appendix B, for $K=2,3$ and 4 extreme profiles.

The main objective of the analysis with this model is to compare the underlying distribution of the membership vectors conditional on generational groups, as a way of assessing differences in the ways of aging between different cohorts. We do this by directly comparing the parameters of these distributions for each generational group $\gamma \in \Gamma$, $\boldsymbol{\alpha}^{\gamma}$, and interpreting them with respect to the common extreme trajectories, defined by the parameters ($\boldsymbol{\beta}_{jk}$).

Figure 6 shows the estimates (posterior means) of the components of the vector $\boldsymbol{\xi}$ for models with $K=2,3$ and 4 extreme profiles, for each cohort. For each



(a) $K = 2$ (b) $K = 3$ (c) $K = 4$

FIG. 6. *Evolution of the parameter vector $\boldsymbol{\xi}$ across different generations for models with $K = 2, 3$ and 4 extreme profiles. The error bars show 95% equal tail posterior credible intervals associated with the kth component of the vector $\boldsymbol{\xi}$.*

generational group, γ_c, the sequence of values of $\xi_k^{\gamma_c}$ are linked with lines. Reading from left to right, these sequences indicate the evolution of the relative weight of the kth component in each cohort as we shift our attention from older to younger cohorts. Posterior estimates of the common extreme profile parameters, $(\boldsymbol{\beta}_{jk})$, are very similar to those computed using the basic TGoM model (see the supplementary material [Manrique-Vallier (2014)] for details), so we can safely refer to them when discussing extreme profiles.

The most salient feature in Figure 6 is the increasing monotonicity of the relative importance of the first component ($k = 1$) in each cohort as we consider younger and younger cohorts, that is, $\xi_1^{\gamma_1} < \xi_1^{\gamma_2} < \cdots < \xi_1^{\gamma_5}$. 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. A likely explanation for this uncertainty is the lack of data for ages past 78 years old in cohort 5.

This trend tells us that, as we consider newer cohorts, their members tend to be increasingly close to profile $k = 1$. This profile corresponds to the healthiest aging progression, with extremely low probability of acquiring disabilities until very advanced ages, as can be observed in Figure 3. Thus, we conclude that younger generations tend to have healthier ways of aging compared to their elders.

6. Discussion. The methods we propose and apply here have several desirable features. First, they produce meaningful and easy-to-interpret summaries of the main temporal trends in the population. In this application, these summaries—the extreme trajectories—isolate typical ways of progressing into disability and allow a simplified analysis of the longitudinal patterns. Second, they allow a simple, but not oversimplified, characterization of the individual heterogeneity in terms of the extreme trajectories. This keeps the extreme profile characterizations simple, while still allowing the representation of complex individual trajectories. Finally, the model's extensions allow comparisons between groups of individuals defined

by given static characteristics. In this application it enables the separation of time-dependent effects that depend on age from those dependent on birth cohort.

The results obtained through the application of our methods to the NLTCS highlight some interesting characteristics of the data and, in general, of the aging process in the U.S. All the models considered here showed that most individuals are close to the "healthy aging" profile ($k = 1$), whose associated extreme trajectories (for the 6 ADLs) describe a practically disability-free life until very late ages (90+). Then we find that profiles with trajectories that specify earlier onsets of disability exhibit progressively less importance in the population. This means that most people could be expected to have a relatively disability-free old age and that very bad aging processes are not so common.

When considering the effect of the birth cohort—estimating simultaneously population-wide extreme profiles together with individual membership conditional on cohort—we find a similar situation. However, different generations have a different membership composition: the relative importance of the "healthy aging" profile ($k = 1$) experiences a monotonic increase when moving from older to younger generations, to the detriment of all the other profiles. Thus, the answer to the question "do younger generations acquire disabilities differently than older ones?" appears to be affirmative. Furthermore, it is so in a positive sense: not only do younger generations acquire disabilities differently, they also acquire them later. These findings are consistent with previous evidence showing a decline in disability obtained from purely cross-sectional analyses [Manton, Corder and Stallard (1997), Manton, Gu and Lamb (2006)], from wave to wave latent class transition analysis [White and Erosheva (2013)], and from latent trajectory analysis [Connor (2006)].

So far declines in disability have been analyzed mostly from wave to wave, either from uncorrelated cross-sectional samples as changes in prevalence [Manton, Corder and Stallard (1997), Manton, Gu and Lamb (2006), Manton, Lamb and Gu (2007)] or from longitudinal analysis as transitions between states [Stallard (2005), White and Erosheva (2013)]. Our approach, in contrast, is not rooted on survey waves, nor does it directly assess changes in prevalence of disability. Instead, it characterizes whole individual life trajectories. It therefore enables direct comparisons across different ways of aging.

An important issue that we have addressed informally here is choosing the number of extreme profiles, K. In Section 5.1.2 we noted that in general the out-of-sample multivariate fit measures improved with model complexity, increasing K, although the improvement was different depending on which multivariate dimension we chose to analyze. We also have observed that the least important profiles in models with $K > 4$ do not reveal informative trajectories. Furthermore, we note that our conclusions do not really depend on an exact number of extreme profiles. Therefore, we evaluate that in this case we do not need to select a "best" model; instead, we have opted for reporting results from several models, with numbers of extreme profiles ranging from $K = 2$ to $K = 4$. Model selection, however, can

be an important issue in other applications. Possible approaches include the use of indexes such as AIC [Akaike (1973)] or BIC [Schwarz (1978)]—or their more computationally convenient counterparts DIC [Spiegelhalter et al. (2002)], AICM or BICM [Raftery et al. (2007)]—although in this case the difficulty in computing the integrated likelihood could make this approach impractical. Another approach is to use Bayesian nonparametric specification that favor sparse representations, such as Dirichlet Process mixtures. Bhattacharya and Dunson (2012) have proposed what is essentially a nonparametric Mixed Membership model for categorical data which could be adapted for this purpose.

As for model limitations, this analysis attributes all variability in the data to a combination of random fluctuation, age effects, cohort effects and Mixed Membership. Thus, it neglects other potential systematic effects, some of which might be important either to capture previously unaccounted variability or simply for better understanding the underlying processes. For instance, it is well known [see, e.g., Ferrucci et al. (1996), Manton (2008)] that men and women follow different aging and mortality processes. One natural way of accounting for nontime-dependent categorical covariates, like gender or race, is to introduce them in the same way we introduced the DOB covariate: as conditioners on the prior distribution of individual membership. If the cells on the contingency table generated by the cross-classification according to the covariates are well populated, we can directly use the TGoM extensions from Section 3.3. If this is not the case, the joint covariate vector can be smoothed using more complex prior specifications, such as those proposed in Bertolet (2008) for the Grade of Membership model.

Another related limitation of these models is that they do not account for mortality. In essence, these models correspond to what Kurland and Heagerty (2005) and Kurland et al. (2009) call an "immortal cohort." This is of particular importance in the present application because patterns of disability are usually tied to patterns of mortality [Connor (2006), Ferrucci et al. (1996), White and Erosheva (2013)]: progression into more severe disability goes together with an increased probability of death. One way of integrating mortality into this framework is to extend the definition of extreme profiles to characterize not only patterns of disability acquisition, but also of survival. Such a joint model could be the topic of a future article.

APPENDIX A: MCMC SAMPLER FOR THE TGOM MODEL

In this appendix we present a Gibbs sampling algorithm for Bayesian estimation of the TGoM model, also described in Manrique-Vallier (2014). Following the discussion at the end of Section 3.1, we construct an algorithm for obtaining samples from the posterior distribution of parameters in the augmented data model in equation (3.6), which after marginalizing $\mathbf{z}$ is equivalent to the TGoM model. This posterior distribution is

$$p(\boldsymbol{\alpha}, \boldsymbol{\beta}, \mathbf{Z}, \mathbf{g}|\mathbf{Y}, \mathbf{Age}) \propto p(\boldsymbol{\alpha}, \mathbf{g}, \boldsymbol{\beta}) \prod_{i=1}^{N} f^{\mathrm{AUG}}(\mathbf{Y}_i, \mathbf{Z}_i|\mathbf{Age}_i, \boldsymbol{\beta}, \mathbf{g}_i), \tag{A.1}$$

which following the detailed specification from Section 3.2 is equivalent to

$$
\begin{aligned}
&p(\boldsymbol{\alpha}, \boldsymbol{\beta}, \mathbf{Z}, \mathbf{g}|\mathbf{Y}, \mathbf{Age}) \\
&\quad \propto \mathrm{Gamma}(\alpha_0|a_\alpha, b_\alpha) \times \mathrm{Dirichlet}(\boldsymbol{\xi}|\mathbf{1}_K) \\
&\quad \times \prod_{i=1}^{N} \mathrm{Dirichlet}(\mathbf{g}_i|\boldsymbol{\alpha}) \times \prod_{j=1}^{J}\prod_{k=1}^{K} N(\beta_{0jk}|\mu_0, \sigma_0^2) \times N(\beta_{1jk}|\mu_1, \sigma_1^2) \\
&\quad \times \prod_{i=1}^{N}\prod_{j=1}^{J}\prod_{t=1}^{T} g_{iz_{ijt}} \frac{\exp(y_{ijt}\beta_{0jz_{ijt}} + y_{ijt}\beta_{1jz_{ijt}}\mathrm{Age}_{it})}{1+\exp(\beta_{0jz_{ijt}} + \beta_{1jz_{ijt}}\mathrm{Age}_{it})},
\end{aligned}
\tag{A.2}
$$

with $\alpha_0 = \sum \alpha_k$ and $\boldsymbol{\xi} = (\alpha_1/\alpha_0, \dots, \alpha_K/\alpha_0)$. Parameters a_α and b_α are shape and inverse scale parameters, respectively.

A Gibbs sampling algorithm for obtaining samples from the joint posterior distribution of $(\boldsymbol{\alpha}, \boldsymbol{\beta}, \mathbf{Z}, \mathbf{g})$ can be constructed as follows:

1. *Sampling from* $\mathbf{Z}$: For every $i \in \{1, \dots, N\}$, $j \in \{1, \dots, J\}$ and $t \in \{1, \dots, T\}$, sample $z_{ijt}|\cdots \sim \mathrm{Discrete}(\{1, \dots, K\}, (p_1, p_2, \dots, p_K))$, with

$$
p_k \propto g_{ik} \frac{\exp[y_{ijt}(\beta_{0jk} + \beta_{1jk}\mathrm{Age}_{it})]}{1+\exp(\beta_{0jk} + \beta_{1jk}\mathrm{Age}_{it})}
$$

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

2. *Sampling from* $\boldsymbol{\beta}_{jk}$: let $\Xi = \{(i,t) : z_{ijt} = k\}$ and assume that $\mu_0 = \mu_1 = 0$. The full joint conditional distribution of $(\beta_{0jk}, \beta_{1jk})$ is

$$
\begin{aligned}
&p(\beta_{0jk}, \beta_{1jk}|\cdots) \\
&\quad \propto \frac{\exp[-(\beta_{1jk}^2/2\sigma_1^2 + \beta_{0jk}^2/2\sigma_0^2) + \beta_{0jk}\sum_\Xi y_{ijt} + \beta_{1jk}\sum_\Xi \mathrm{Age}_{it}y_{ijt}]}{\prod_\Xi[1+\exp(\beta_{0jk} + \beta_{0jk}\mathrm{Age}_{it})]}.
\end{aligned}
$$

To sample from this distribution, we use a random walk Metropolis step:

(a) Sample proposal values $\beta_{0jk}^* \sim N(\beta_{0jk}, \sigma_{\beta 0}^2)$ and $\beta_{1jk}^* \sim N(\beta_{1jk}, \sigma_{\beta 1}^2)$, where $\sigma_{\beta 0}^2$ and $\sigma_{\beta 1}^2$ are tuning parameters.

(b) With probability

$$
\begin{aligned}
r_M = \min\Bigg\{1, \prod_\Xi \Bigg[\frac{1+\exp[\beta_{0jk} + \beta_{0jk}\mathrm{Age}_{it}]}{1+\exp[\beta_{0jk}^* + \beta_{0jk}^*\mathrm{Age}_{it}]}\Bigg] \\
\times \exp\Bigg[-\frac{\beta_{0jk}^{*2} - \beta_{0jk}^2}{2\sigma_0^2} + (\beta_{0jk}^* - \beta_{0jk})\sum_\Xi y_{ijt}\Bigg] \\
\times \exp\Bigg[-\frac{\beta_{1jk}^{*2} - \beta_{1jk}^2}{2\sigma_1^2} + (\beta_{1jk}^* - \beta_{1jk})\sum_\Xi y_{ijt}\mathrm{Age}_{it}\Bigg]\Bigg\},
\end{aligned}
\tag{A.3}
$$

make $(\beta_{0jk}, \beta_{1jk}) = (\beta_{0jk}^*, \beta_{1jk}^*)$. Otherwise keep the current value.

3. *Sampling from* $\mathbf{g}_i$:

$$\mathbf{g}_i|\cdots \overset{\text{indep}}{\sim} \text{Dirichlet}\Big(\alpha_1 + \sum_{j,t} I(z_{ijt}=1), \dots, \alpha_K + \sum_{j,t} I(z_{ijt}=K)\Big).$$

4. *Sampling from* $\boldsymbol{\alpha}$: The full conditional distribution of $\boldsymbol{\alpha}$,

$$p(\boldsymbol{\alpha}|\cdots) \propto \alpha_0^{a_\alpha - 1} e^{-\alpha_0 b_\alpha} \times \left[\frac{\Gamma(\alpha_0)}{\prod_{k=1}^K \Gamma(\alpha_k)}\right]^N \prod_{k=1}^K \left[\prod_{i=1}^N g_{ik}\right]^{\alpha_k},$$

does not have any recognizable form. We use a Metropolis–Hastings step similar to Manrique-Vallier and Fienberg (2008):

(a) Obtain the proposal

$$\boldsymbol{\alpha}^* = (\alpha_1^*, \alpha_2^*, \dots, \alpha_K^*) \qquad \text{with } \alpha_k^* \overset{\text{indep}}{\sim} \text{lognormal}(\log \alpha_k, \sigma^2).$$

(b) Let $\alpha_0^* = \sum_{k=1}^K \alpha_k^*$. With probability

$$r = \min\left\{1, e^{-a_\alpha(\alpha_0^* - \alpha_0)} \left(\frac{\alpha_0^*}{\alpha_0}\right)^{b_\alpha - 1} \left(\prod_{k=1}^K \frac{\alpha_k^*}{\alpha_k}\right) \times \left[\frac{\Gamma(\alpha_0^*)}{\Gamma(\alpha_0)} \prod_{k=1}^K \frac{\Gamma(\alpha_k)}{\Gamma(\alpha_k^*)}\right]^N \prod_{k=1}^K \left(\prod_{i=1}^N g_{ik}\right)^{\alpha_k^* - \alpha_k}\right\},$$

make $\boldsymbol{\alpha} = \boldsymbol{\alpha}^*$. Otherwise keep the current value. Obtain $(\alpha_0, \boldsymbol{\xi})$ by making $\alpha_0 = \sum_{k=1}^K \alpha_k$ and $\xi_k = \alpha_k/\alpha_0$, for all $k = 1, \dots, K$.

APPENDIX B: FITTING THE GENERATIONAL EXTENSION

The only difference between the posterior distributions of the basic and the extended TGoM models is the distribution of $\mathbf{g}_i|\boldsymbol{\alpha}$. Thus, we only have to adapt steps 3 and 4 in the previous algorithm by replacing $\prod_{i=1}^N p(\mathbf{g}_i|\boldsymbol{\alpha})$ with

$$\prod_{i=1}^N p(\mathbf{g}_i|\boldsymbol{\alpha}(\text{DOB}_i)) = \prod_{\gamma\in\Gamma} \prod_{i=1}^N [p(\mathbf{g}_i|\boldsymbol{\alpha}^\gamma)]^{I(\text{DOB}_i \in \gamma)}. \tag{B.1}$$

Let $\gamma_i \in \Gamma$ be the unique interval from the partition such that $\text{DOB}_i \in \gamma_i$. We obtain an MCMC sampler for this model by modifying steps 3 and 4 from the algorithm in Appendix A with the following:

3′. *Sampling from* $\mathbf{g}_i$:

$$\mathbf{g}_i|\cdots \overset{\text{indep}}{\sim} \text{Dirichlet}\Big(\alpha_1^{\gamma_i} + \sum_{j,t} I(z_{ijt}=1), \dots, \alpha_K^{\gamma_i} + \sum_{j,t} I(z_{ijt}=K)\Big).$$

4′. *Sampling from* $\boldsymbol{\alpha}$: let $\Xi_\gamma = \{i : \gamma_i = \gamma\}$. The full conditional distribution of $\boldsymbol{\alpha}^\gamma$ is

$$p(\boldsymbol{\alpha}^\gamma \mid \cdots) \propto (\alpha_0^\gamma)^{a_\alpha - 1} e^{-\alpha_0^\gamma b_\alpha} \times \left[\frac{\Gamma(\alpha_0^\gamma)}{\prod_{k=1}^K \Gamma(\alpha_k^\gamma)}\right]^{\#(\Xi_\gamma)} \prod_{k=1}^K \left[\prod_{\Xi_\gamma} g_{ik}\right]^{\alpha_k^\gamma},$$

where $\#(\Xi_\gamma)$ is the number of elements in the set Ξ_γ.

This expression is similar to (B.1) in Appendix A. We thus adapt the procedure by replacing r in step 4 of the algorithm with

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

Acknowledgments. The author would like to thank Professor Steve Fienberg for his guidance and encouragement while conducting this investigation, Professor Jerry Reiter for his help and valuable suggestions during the review process, and three anonymous referees and one Associate Editor for their thorough critique and recommendations. This work is partially based on material from the Ph.D. thesis of the author in the Department of Statistics at Carnegie Mellon University. For a closely related treatment of the model and data analysis see also Manrique-Vallier (2014).

SUPPLEMENTARY MATERIAL

Supplement to "Longitudinal Mixed Membership trajectory models for disability survey data" (DOI: 10.1214/14-AOAS769SUPP; .pdf). Estimation using TGoM models with piecewise constant trajectories and tables with posterior estimates for all the fitted models.

REFERENCES

AIROLDI, E. M., FIENBERG, S. E., JOUTARD, C. and LOVE, T. M. (2007). Discovering latent patterns with hierarchical Bayesian mixed-membership models. In *Data Mining Patterns*: *New Methods and Applications* (P. Poncelet, F. Masseglia and M. Teisseire, eds.) 240–275. Idea Group Inc., Hershey, PA.

AIROLDI, E. M., BLEI, D. M., FIENBERG, S. E. and XING, E. P. (2008). Mixed membership stochastic blockmodels. *J. Mach. Learn. Res.* **9** 1981–2014.

AIROLDI, E. M., EROSHEVA, E. A., FIENBERG, S. E., JOUTARD, C., LOVE, T. and SHRINGARPURE, S. (2010). Reconceptualizing the classification of PNAS articles. *Proc. Natl. Acad. Sci. USA* **107** 20899–20904.

AKAIKE, H. (1973). Information theory and an extension of the maximum likelihood principle. In *Second International Symposium on Information Theory* 267–281. Akademiai Kiado, Budapest.

BERTOLET, M. (2008). To weight or not to weight? Incorporating sampling designs into model-based analyses. Ph.D. thesis, Dept. Statistics, Carnegie Mellon Univ., Pittsburgh, PA.

BHATTACHARYA, A. and DUNSON, D. B. (2012). Simplex factor models for multivariate unordered categorical data. *J. Amer. Statist. Assoc.* **107** 362–377. MR2949366

CLARK, R. F. (1998). An introduction to the National Long-Term Care Survey. Office of Disability, Aging, and Long-Term Care Policy with the U.S. Dept. Health and Human Services. Available at http://aspe.hhs.gov/daltcp/reports/nltcssu2.htm.

CONNOR, J. T. (2006). Multivariate mixture models to describe longitudinal patterns of frailty in American seniors. Ph.D. thesis, Dept. Statistics & H. John Heinz III School of Public Policy & Management, Carnegie Mellon Univ., Pittsburgh, PA.

CONNOR, J. T., FIENBERG, S. E., EROSHEVA, E. A. and WHITE, T. (2006). Towards a restructuring of the National Long Term Care Survey: A longitudinal perspective. Prepared for presentation at an Expert Panel Meeting on the National Long Term Care Survey, Committee on National Statistics, National Research Council.

CORDER, L. S. and MANTON, K. G. (1991). National surveys and the health and functioning of the elderly: The effects of design and content. *J. Amer. Statist. Assoc.* **86** 513–525.

EROSHEVA, E. A. (2002). Grade of membership and latent structures with application to disability survey data. Ph.D. thesis, Dept. Statistics, Carnegie Mellon Univ., Pittsburgh, PA.

EROSHEVA, E. A. and FIENBERG, S. E. (2005). Bayesian mixed membership models for soft clustering and classification. In *Classification—The Ubiquitous Challenge* (C. Weihs and W. Gaul, eds.) 11–26. Springer, Berlin.

EROSHEVA, E. A., FIENBERG, S. E. and JOUTARD, C. (2007). Describing disability through individual-level mixture models for multivariate binary data. *Ann. Appl. Stat.* **1** 502–537. MR2415745

EROSHEVA, E. A., FIENBERG, S. E. and LAFFERTY, J. D. (2004). Mixed-membership models of scientific publications. *Proc. Natl. Acad. Sci. USA* **101** 5220–5227.

FERRUCCI, L., GURALNIK, J. M., SIMONSICK, E., SALIVE, M. E., 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.

GOODMAN, L. A. (1974). Exploratory latent structure analysis using both identifiable and unidentifiable models. *Biometrika* **61** 215–231. MR0370936

HASTIE, T., TIBSHIRANI, R. and FRIEDMAN, J. (2009). *The Elements of Statistical Learning*: *Data Mining*, *Inference*, *and Prediction*, 2nd ed. Springer, New York. MR2722294

JASRA, A., HOLMES, C. C. and STEPHENS, D. A. (2005). Markov chain Monte Carlo methods and the label switching problem in Bayesian mixture modeling. *Statist. Sci.* **20** 50–67. MR2182987

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. F. and HEAGERTY, P. J. (2005). Directly parameterized regression conditioning on being alive: Analysis of longitudinal data truncated by deaths. *Biostatistics* **6** 241–258.

KURLAND, B. F., JOHNSON, L. L., EGLESTON, B. L. and DIEHR, P. H. (2009). Longitudinal data with follow-up truncated by death: Match the analysis method to research aims. *Statist. Sci.* **24** 211–222. MR2655850

MANRIQUE-VALLIER, D. (2014). Mixed membership trajectory models. In *Handbook on Mixed Membership Models* (E. M. Airoldi, D. M. Blei, E. A. Erosheva and S. E. Fienberg, eds.). Chapman & Hall/CRC, London.

MANRIQUE-VALLIER, D. (2014). Supplement to "Longitudinal Mixed Membership trajectory models for disability survey data." DOI:10.1214/14-AOAS769SUPP.

MANRIQUE-VALLIER, D. and FIENBERG, S. E. (2008). Population size estimation using individual level mixture models. *Biom. J.* **50** 1051–1063. MR2649394

MANTON, K. G. (2008). Recent declines in chronic disability in the elderly U.S. population: Risk factors and future dynamics. *Annu. Rev. Public Health* **29** 91–113.

MANTON, K. G., CORDER, L. and STALLARD, E. (1997). Chronic disability trends in elderly United States populations: 1982–1994. *Proc. Natl. Acad. Sci. USA* **94** 2593–2598.

MANTON, K. G., GU, X. L. and LAMB, V. L. (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. *Proc. Natl. Acad. Sci. USA* **103** 18374.

MANTON, K. G., LAMB, V. L. and GU, X. (2007). Medicare cost effects of recent US disability trends in the elderly future implications. *J. Aging Health* **19** 359–381.

MANTON, K. G., STALLARD, E. and WOODBURY, M. A. (1991). A multivariate event history model based upon fuzzy states: Estimation from longitudinal surveys with informative nonresponse. *J. Off. Stat.* **7** 261–293.

NAGIN, D. S. (1999). Analyzing developmental trajectories: A semiparametric, group-based approach. *Psychol. Methods* **4** 139–157.

RAFTERY, A. E., NEWTON, M. A., SATAGOPAN, J. M. and KRIVITSKY, P. N. (2007). Estimating the integrated likelihood via posterior simulation using the harmonic mean identity. In *Bayesian Statistics* (J. M. Bernardo, M. J. Bayarri, J. O. Berger, A. P. Dawid, D. Heckerman, A. F. M. Smith and M. West, eds.). *Oxford Sci. Publ.* **8** 371–416. Oxford Univ. Press, Oxford. MR2433201

SCHWARZ, G. (1978). Estimating the dimension of a model. *Ann. Statist.* **6** 461–464. MR0468014

SPIEGELHALTER, D. J., BEST, N. G., CARLIN, B. P. and VAN DER LINDE, A. (2002). Bayesian measures of model complexity and fit. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **64** 583–639. MR1979380

STALLARD, E. (2005). Trajectories of morbidity, disability, and mortality among the US elderly population: Evidence from the 1984–1999 NLTCS. *N. Amer. Actuar. J.* **11** 16–53.

TANNER, M. A. (1996). *Tools for Statistical Inference*: *Methods for the Exploration of Posterior Distributions and Likelihood Functions*, 3rd ed. Springer, New York. MR1396311

WHITE, T. A. and EROSHEVA, E. A. (2013). Using group-based latent class transition models to analyze chronic disability data from the National Long-Term Care Survey 1984–2004. *Stat. Med.* **32** 3569–3589. MR3092251

WOODBURY, M. A., CLIVE, J. and GARSON, A. JR. (1978). Mathematical typology: A grade of membership technique for obtaining disease definition. *Comput. Biomed. Res.* **11** 277–98.

XING, E. P., FU, W. and SONG, L. (2010). A state-space mixed membership blockmodel for dynamic network tomography. *Ann. Appl. Stat.* **4** 535–566. MR2758639

DEPARTMENT OF STATISTICS
INDIANA UNIVERSITY
BLOOMINGTON, INDIANA 47408
USA
E-MAIL: dmanriqu@indiana.edu