



HHS Public Access

Author manuscript

J Comput Graph Stat. Author manuscript; available in PMC 2017 January 05.

Published in final edited form as:

J Comput Graph Stat. 2010 ; 19(3): 746–765. doi:10.1198/jcgs.2010.09015.

Multivariate Discrete Hidden Markov Models for Domain-Based Measurements and Assessment of Risk Factors in Child Development

Qiang Zhang [Research Associate],

Department of Biostatistical Sciences, Wake Forest University School of Medicine, Winston-Salem, NC 27159.

Alison Snow Jones [Associate Professor],

Business and Public Health Partnerships, Drexel University School of Public Health, Philadelphia, PA 19102.

Frank Rijmen [Senior Research Scientist], and

Educational Testing Service, Princeton, NJ 08541.

Edward H. Ip [Professor]

Department of Biostatistical Sciences and Department of Social Sciences and Health Policy, Wake Forest University School of Medicine, Winston-Salem, NC 27159 (eip@wfubmc.edu).

Abstract

Many studies in the social and behavioral sciences involve multivariate discrete measurements, which are often characterized by the presence of an underlying individual trait, the existence of clusters such as domains of measurements, and the availability of multiple waves of cohort data. Motivated by an application in child development, we propose a class of extended multivariate discrete hidden Markov models for analyzing domain-based measurements of cognition and behavior. A random effects model is used to capture the long-term trait. Additionally, we develop a model selection criterion based on the Bayes factor for the extended hidden Markov model. The National Longitudinal Survey of Youth (NLSY) is used to illustrate the methods. Supplementary technical details and computer codes are available online.

Keywords

Junction tree; Mixed effects; National Longitudinal Survey of Youth

1. INTRODUCTION

Empirical studies in social and behavioral research often involve multivariate discrete data. Dichotomized responses from multiple items on a survey, which are indicators of the

SUPPLEMENTAL MATERIALS

Appendix: Technical details on GLM estimation, the proof of Lemma 2, and instructions to use the computer codes. (appendix.pdf)

Computer codes: MATLAB-package containing codes to fit Models I to IV in Section 4.1 with instructions appearing in the online Appendix. The package also contains a subset of the NLSY dataset used in this article. (MHMM.zip)

presence of a multitude of behaviors (or the lack thereof), are routinely included in such studies. The past decade has also witnessed, for good reasons, a tremendous growth of cohort data in the social sciences (Frees 2004, p. 4; Singer and Willet 2003). Unlike cross-sectional data, longitudinal data collected from a cohort are valuable for informing the researcher about the dynamics of behavior and changing attitudes and perceptions in a population. As an example, the field of child development traditionally has sought to describe universals of development and to investigate mechanisms of development through the use of small-scale, cross-sectional experiments. However, emphasis has recently been shifting from this kind of so-called variable-centered study to person-centered longitudinal studies, which include large-scale, long-term projects such as the National Longitudinal Survey of Youth (NLSY, Chase-Lansdale et al. 1991). These studies often aim to identify subgroups of differential risks, and to delineate risk and protective factors that explain how development proceeds differently in different subgroups (Laursen and Hoff 2006). Indeed, in the NLSY, extensive cognitive and linguistic outcome measures on a national sample of children are being collected over an extended period of time, and these measures can be related to external variables such as socioeconomic status (SES) and maternal health, both mental and physical.

The discrete hidden Markov model (HMM, Rabiner 1989; MacDonald and Zucchini 1997) is a natural way for modeling discrete multiple outcomes (also referred to interchangeably as “responses” in this article) collected on individuals or social units over time. The HMM assumes the presence of two parallel stochastic processes: an observed sequence of outcomes and a Markovian sequence of underlying latent states that drives the observed stochastic changes. Observed outcome variables over time are assumed to be conditionally independent given latent state. Furthermore, when more than one outcome variable is involved, conditional independence between multiple outcomes given latent state is also assumed.

Although the multivariate HMM has many features necessary for successful person-centered longitudinal studies, several modeling and computational limitations exist. First, while the conditional independence assumption greatly simplifies computation, the assumption may not be realistic in many social science applications. In the NLSY dataset, a child-developmental profile may contain clearly identifiable domains such as problematic behaviors and cognitive development. It is unlikely that measurements within the same domain (e.g., oral reading and reading comprehension within the domain of cognitive development) would satisfy the conditionally independent assumption given latent state.

Second, in person-centered longitudinal studies, there often exist discernible individual stable traits that “drive” an individual’s responses both across domains and across time points. In a longitudinal study of domestic violence, Snow Jones et al. (2010) found evidence that a relatively stable individual trait of violent behavior exists among batterers, despite the fact that their behaviors may cycle back and forth from overtly violent behavioral states to relatively nonabusive behavioral states. Statistically, the stable trait can be modeled as an individual-specific random effect that induces a level of conditional correlation across outcomes and time points given latent state. In other words, the conditional independence

assumption in HMM is relaxed to allow for heterogeneity among individuals within a specific latent state.

A third challenge in person-centered longitudinal studies is that conventional indexes for model selection in HMM, such as the AIC and BIC, may be invalid because observations across time points are generally not independent. For example, in BIC, the sample size needs to be adjusted to account for the correlation between observations across time.

This article describes our effort to resolve the aforementioned limitations of the HMM, especially in the context of social and behavioral data. We discuss (1) a domain-based model for handling the existence of domains of items within the data, (2) a random effects model for capturing the individual-specific stable trait, and (3) a Bayes factor-based goodness-of-fit index, which we call the modified BIC, that takes into account the correlation between observations across time points. Related computational methods for implementing the three solutions will be discussed.

Several works in the literature of HMM are especially pertinent to this article. In an application of the HMM to rainfall data, Zucchini and Guttorp (1991) suggested modifications to relax the conditional independence assumption across their multisite precipitation model because sites in close proximity may exhibit similar rainfall patterns. Hughes, Guttorp, and Charles (1999) described a nonhomogeneous HMM for precipitation. Scott, James, and Sugar (2005) described an HMM for multiple, continuous observed data that are assumed to follow a multivariate t -distribution. Altman (2007) proposed mixed-effects HMMs and applied the model to lesion-count data in a two-state Poisson model. The author also proposed using a quasi-Newton method for HMM and either numerical integration or Monte Carlo methods for handling the random effects. Rijmen, Vansteelandt, and De Boeck (2008) proposed, from an item-response theory (IRT, Lord and Novick 1968) perspective, HMMs for analyzing psychological data. Rijmen et al. (2008) applied the HMM to biomedical data collected from patients with brain tumor. A structurally identical family of models—the latent Markov models—has also been extensively studied in the social science literature. Some recent applications in social and economic sciences include the work of Bartolucci, Pennoni, and Francis (2007), Paas, Bijmolt, and Vermunt (2007), and Ip et al. (2010). Dias and Vermunt (2007) discussed model selection issues in latent Markov and mixture models. For a history of development of latent Markov models, see the works by Wiggins (1973), Humphreys (1998), Vermunt, Langeheine, and Bockenholt (1999), and the references therein.

The remainder of this article is organized as follows. We first describe the basic multivariate discrete HMM model and methods of estimating the models. Then we describe the aforementioned methods (1)–(3) for solving the several limitations of the HMM. Subsequently, we describe an application using NLSY data to illustrate the proposed methods.

2. MODEL

2.1 The Basic Discrete Hidden Markov Model

Let y_{ijt} denote the discrete response of subject i at occasion t on outcome j , $i = 1, \dots, N$; $j = 1, \dots, J$; $t = 1, \dots, T$. We assume that each outcome indicates the level of response k , $k = 1, \dots, K_j$ from an individual as measured by an “item,” such as a survey question. Specifically, $\mathbf{y}_{it} = (y_{it1}, \dots, y_{itJ})'$ denotes the response vector of subject i at occasion t , and $\mathbf{y}_i = (\mathbf{y}'_{i1}, \dots, \mathbf{y}'_{iT})'$ represents the complete response pattern of subject i . Additionally, $z_{it} = 1, \dots, S$ denotes the categorical (unobserved) latent state of subject i at occasion t . Thus, $\mathbf{z}_i = (z_{i1}, \dots, z_{iT})'$ is the state sequence of subject i through the latent space over time. A first-order Markov chain for the latent variable assumes that the latent state at occasion $t + 1$ depends only upon the latent state at occasion t ; that is, for any i , $p(z_{it+1} | z_{i1}, \dots, z_{iT}) = p(z_{it+1} | z_{it})$. Accordingly, the basic hidden Markov model contains the following three sets of parameters:

1. $\boldsymbol{\tau} = (\tau_{rs})$, the $S \times S$ matrix of time-homogeneous transition probabilities between latent states, where for any i , $\tau_{rs} = p(z_{it} = s | z_{it-1} = r)$, $t = 2, \dots, T$.
2. $\boldsymbol{\alpha}_1 = (\alpha_{11}, \dots, \alpha_{1S})'$, the $S \times 1$ vector of marginal probabilities of the states at occasion 1, $\alpha_{1s} = p(z_{i1} = s)$. The vector $\boldsymbol{\alpha}_t = (\alpha_{t1}, \dots, \alpha_{tS})'$, $t = 2, \dots, T$, can be recursively derived by the formula $\boldsymbol{\alpha}'_t = \boldsymbol{\alpha}'_{t-1} \boldsymbol{\tau}$.
3. $\boldsymbol{\pi} = (\pi_{jks})$, the $J \times S \times K_j$ array of state-conditional response probabilities, where k is the index of outcome category, $k = 1, \dots, K_j$ and for any i, t , $\pi_{jks} = \Pr(y_{ijt} = k | z_{it} = s)$, where $\sum_{k=1}^{K_j} \pi_{jks} = 1$.

Assuming conditional independence between responses given the latent state, the marginal probability of a response pattern \mathbf{y}_i is given by

$$p(\mathbf{y}_i) = \sum_{\mathcal{Z}} p(\mathbf{z}_i) \prod_{t=1}^T p(\mathbf{y}_{it} | z_{it}), \quad (2.1)$$

with the summation over the entire latent state space \mathcal{Z} of size S^T , and

$$p(\mathbf{z}_i) = \alpha_{1z_{i1}} \prod_{t=2}^T \tau_{z_{it-1}z_{it}}. \quad (2.2)$$

Furthermore, we write out the likelihood equation using the local independence assumption given a latent state:

$$p(\mathbf{y}_{it} | z_{it} = s) = \prod_{j=1}^J \prod_{k=1}^{K_j} \pi_{jks}^{\delta(y_{itj}, k)}, \quad (2.3)$$

where $\delta(y_{itj}, k) = 1$ if $y_{itj} = k$, and 0 otherwise.

2.2 Estimation of the Basic HMM

The EM algorithm (Dempster, Laird, and Rubin 1977) is used to evaluate parameters for the basic HMM. For historical reasons, a version of the EM algorithm used to estimate HMM is also known as the Baum–Welch algorithm (Baum and Petrie 1966). Briefly, in the EM terminology, the latent state is treated as “missing data.” The objective here is to maximize the conditional expected log-likelihood of the complete data, which include the observed data \mathbf{y} and the latent states \mathbf{z} , given \mathbf{y} and some provisional estimates for the parameters, where \mathbf{y} and \mathbf{z} , respectively, denote the vector of the collection of data (\mathbf{y}_i) and (\mathbf{z}_i).

Denote the entire parameter set defined in (2.1)–(2.2) by β , and let $\beta^{(n)}$ be defined as the estimate at the n th iteration of the EM algorithm. Furthermore, let

$$\beta^{(n+1)} = \arg \max_{\beta} Q(\beta, \beta^{(n)}),$$

where Q is the conditional expected value of the complete data log-likelihood function given observed data \mathbf{y} :

$$Q(\beta, \beta^{(n)}) = E[\log(p(\mathbf{y}, \mathbf{z}|\beta)) | \mathbf{y}, \beta^{(n)}]. \quad (2.4)$$

With discrete latent state variables \mathbf{z} , we have

$$Q(\beta, \beta^{(n)}) = \sum_{\mathcal{Z}} p(\mathbf{z} | \mathbf{y}, \beta^{(n)}) \log p(\mathbf{y}, \mathbf{z} | \beta). \quad (2.5)$$

Assuming that observations between individuals are independent, it follows that

$$Q(\beta, \beta^{(n)}) = \sum_i \sum_{\mathcal{Z}} p(\mathbf{z}_i | \mathbf{y}_i, \beta^{(n)}) \log p(\mathbf{y}_i, \mathbf{z}_i | \beta). \quad (2.6)$$

The posterior probabilities, $p(\mathbf{z}_i | \mathbf{y}_i, \beta^{(n)})$, can be derived via the Bayes theorem in the so-called E-step, which we shall describe in Section 3.2. Updating parameter values can be accomplished through the maximization of (2.6) with respect to β in the so-called M-step of the EM algorithm. The algorithm alternates between the E-step and the M-step until the absolute change in the objective function is smaller than a predetermined threshold value.

Specifically, the joint distribution of the complete data of subject i in (2.6) is

$$p(\mathbf{y}_i, \mathbf{z}_i | \beta) = \alpha_{1z_1} \left(\prod_{t=2}^T \tau_{z_{it-1}z_{it}} \right) \prod_{t=1}^T \prod_{j=1}^J \prod_{k=1}^{K_j} \pi_{j sk}^{\delta(y_{itj}, k)}. \quad (2.7)$$

One appealing feature about (2.7) is the separability of three sets of parameters, which can be optimized individually. By taking the logarithm on (2.7), we rewrite the objective function in (2.6):

$$\begin{aligned}
 Q(\beta, \beta^{(n)}) = & \sum_i \sum_{\mathcal{Z}} p(\mathbf{z}_i | \mathbf{y}_i, \beta^{(n)}) \log \alpha_{1z_{i1}} \\
 & + \sum_i \sum_{\mathcal{Z}} p(\mathbf{z}_i | \mathbf{y}_i, \beta^{(n)}) \sum_{t=2}^T \log \tau_{z_{it-1}z_{it}} \\
 & + \sum_i \sum_{\mathcal{Z}} p(\mathbf{z}_i | \mathbf{y}_i, \beta^{(n)}) \sum_t \sum_j \sum_k \delta(y_{itj}, k) \log \pi_{jsk}.
 \end{aligned} \tag{2.8}$$

We further define two sets of subject-specific HMM “pseudo-counts” as follows:

1. $\tilde{\tau}_{it}^{(n)} = (\tilde{\tau}_{itrs}^{(n)})$, where $\tilde{\tau}_{it}^{(n)}$ is an $S \times S$ matrix and for each i , the $(r-s)$ th entry is

$$\tilde{\tau}_{itrs}^{(n)} = p(z_{it} = s, z_{it-1} = r | \mathbf{y}_i, \beta^{(n)}), \quad t = 2, \dots, T. \tag{2.9}$$

2. $\tilde{\alpha}_{i1}^{(n)} = (\tilde{\alpha}_{i11}^{(n)}, \dots, \tilde{\alpha}_{i1S}^{(n)})'$, and $\tilde{\alpha}_{i1s}^{(n)} = p(z_{i1} = s | \mathbf{y}_i, \beta^{(n)})$. The vector $\tilde{\alpha}_{it}^{(n)} = (\tilde{\alpha}_{it1}^{(n)}, \dots, \tilde{\alpha}_{itS}^{(n)})'$, can be recursively derived by the formula $\tilde{\alpha}_{it}^{(n)} = \tilde{\alpha}_{it-1}^{(n)} \tilde{\tau}_{it}^{(n)}$.

Using the notation above, each of the three terms over the right side of (2.8) can be expressed as a multinomial log-likelihood of the form

$$\sum_i \sum_{\mathbf{I}} \mathcal{Y}_{i\mathbf{I}} \log(\psi_{\mathbf{I}}), \tag{2.10}$$

where \mathbf{I} is a vector of summation indices, $\mathcal{Y}_{i\mathbf{I}}$ are the “pseudo-counts,” which may take non-integer values, and $\psi_{\mathbf{I}}$ are the model-based estimated probabilities. Specifically, in the first row of (2.8), $\mathbf{I} = \{s | s = 1, \dots, S\}$, $\mathcal{Y}_{i\mathbf{I}} = \tilde{\alpha}_{i1s}^{(n)}$, and $\psi_{\mathbf{I}} = \alpha_{1s}$; in the second row of (2.8), $\mathbf{I} = \{(r, s) | r = 1, \dots, S, s = 1, \dots, S\}$, $\mathcal{Y}_{i\mathbf{I}} = \sum_{t=2}^T \tilde{\tau}_{itrs}^{(n)}$, and $\psi_{\mathbf{I}} = \tau_{rs}$; and in the third row of (2.8), $\mathbf{I} = \{(j, s, k) | j = 1, \dots, J, s = 1, \dots, S, k = 1, \dots, K_j\}$, $\mathcal{Y}_{i\mathbf{I}} = \sum_t \tilde{\alpha}_{its}^{(n)} \delta(y_{itj}, k)$, and $\psi_{\mathbf{I}} = \pi_{jsk}$.

In (2.10), the pseudo-counts $\tilde{\alpha}_{i1s}^{(n)}$, $\tilde{\tau}_{itrs}^{(n)}$ and $\tilde{\alpha}_{its}^{(n)} \delta(y_{itj}, k)$ can be treated as observed responses in a cell of a multinomial distribution, and α_{1s} , τ_{rs} , and π_{jsk} as the cell probability parameters to be estimated. The following two conditions are satisfied for each of the three components: (1) the expected value of the pseudo-count equals the corresponding parameter, (2) the probability parameter entries sum to 1. For example, for the marginal component, for any i , $E(\alpha_{i1s}^{(n)}) = \alpha_{1s}$, and $\sum_{s=1}^S \alpha_{1s} = 1$. The M-step thus involves applying the sum-to-1

constraints to each of α_{1s} , τ_{rs} , and π_{jsk} , adding a Lagrange multiplier to each term in (2.8), and taking the derivative of each term. The updating equation for maximum likelihood estimates (MLEs) of α_{1s} , τ_{rs} , and π_{jsk} are the solutions of the re-estimation formulas of the basic HMM (Rabiner 1989):

$$\alpha_{1s} = \frac{\sum_{i=1}^N p(z_{i1}=s | \mathbf{y}_i, \beta^{(n)})}{N} = \frac{\sum_{i=1}^N \tilde{\alpha}_{i1s}^{(n)}}{N}, \quad (2.11)$$

$$\tau_{rs} = \frac{\sum_{i=1}^N \sum_{t=2}^T p(z_{it}=s, z_{it-1}=r | \mathbf{y}_i, \beta^{(n)})}{\sum_{i=1}^N \sum_{t=2}^T p(z_{it-1}=r | \mathbf{y}_i, \beta^{(n)})} = \frac{\sum_{i=1}^N \sum_{t=2}^T \tilde{\tau}_{itrs}^{(n)}}{\sum_{i=1}^N \sum_{t=1}^{T-1} \tilde{\alpha}_{itr}^{(n)}}, \quad (2.12)$$

$$\pi_{jsk} = \frac{\sum_{i=1}^N \sum_{t=1}^T p(y_{itj}=k, z_{it}=s | \mathbf{y}_i, \beta^{(n)})}{\sum_{i=1}^N \sum_{t=1}^T p(z_{it}=s | \mathbf{y}_i, \beta^{(n)})} = \frac{\sum_{i=1}^N \sum_{t=1}^T \tilde{\alpha}_{its}^{(n)} \delta(y_{itj}, k)}{\sum_{i=1}^N \sum_{t=1}^T \tilde{\alpha}_{its}^{(n)}}. \quad (2.13)$$

Stated in various forms, the above expression of the conditional expected complete log-likelihood function as multinomial log-likelihood and the estimation formula for the parameters are not new and well known in the literature (e.g., Cowell et al. 1999, p. 202). Our point of stating the result in the above form is to provide the formalism for the subsequent inference procedure when the HMM is extended.

3. EXTENDED MULTIVARIATE DISCRETE HMM

To fix notation, we denote item j within domain d by $d(j)$, $d = 1, \dots, D$, and the predictive p -vector of the individual long-term trait by \mathbf{x}_j . For example, the second item ($j = 2$) measuring anxiety in the domain of problematic behavior ($d = 1$) is subscripted as 1(2). On the other hand, \mathbf{x}_j may contain predictors such as family background. We denote the individual-specific conditional probability $\Pr(y_{itj} = k | z_{it} = s)$ by π_{ijsk} . Note that an index i has been added to the subscript for $\boldsymbol{\pi}$. As we shall see, by extending the HMM to include individual-specific effects, the conditional probabilities given latent states would be functions of i . For discrete data, we use an appropriate link function for transforming the state-conditional response probabilities $\boldsymbol{\pi}$. For the sake of illustration, we only describe the model for ordinal outcomes, although models for binary and unordered categorical responses can be constructed in a similar way. A modified cumulative logit transform (Fahrmeir and Tutz 1994) for the response solicited from item j can be written as

$$h_1(\zeta_{ijs1}) = \log \left(\frac{\zeta_{ijs1}}{1 - \zeta_{ijs1}} \right), \quad (3.1)$$

$$h_k(\zeta_{ij sk}, \zeta_{ijs(k-1)}) = \log \left[\log \left(\frac{\zeta_{ij sk}}{1 - \zeta_{ij sk}} \right) - \log \left(\frac{\zeta_{ijs(k-1)}}{1 - \zeta_{ijs(k-1)}} \right) \right], \quad k=2, \dots, K_j,$$

where the cumulative probability $\zeta_{ij sk}$ is given by

$$\zeta_{ij sk} = \Pr(y_{itj} \leq k | z_{it} = s) = \pi_{ijs1} + \dots + \pi_{ij sk}, \quad (3.2)$$

with $i = 1, \dots, N, j = 1, \dots, J, s = 1, \dots, S$, and $k = 1, \dots, K_j$, and subject to the constraint $\sum_{k=1}^{K_j} \pi_{ij sk} = 1$. By requiring that the difference in logits in (3.1) be positive, the log–log transform enforces the order of the responses such that the cumulative logit of a higher category is always larger than that of a lower category. We shall refer to the collection of link functions (h_1, \dots, h_{K_j-1}) simply as $h(\cdot)$.

We study the following ordinal cumulative logistic models:

(A) Domain-based HMM:

$$h_1(\zeta_{ijs1}) = \phi_{j1} + \xi_{d(j)s1} \quad \text{and} \quad h_k(\zeta_{ij sk}, \zeta_{ijs(k-1)}) = \phi_{jk} + \xi_{d(j)sk}, \quad (3.3)$$

where $k = 2, \dots, K_j - 1; s = 1, \dots, S - 1$. No coefficients ϕ, ξ are defined for the last category K_j , and $\pi_{ijsK_j} = 1 - \sum_{k=1}^{K_j-1} \pi_{ij sk}$. To parameterize (3.3) into a full-rank model, no coefficient is estimated for $\xi_{d(j)sk}$ at the last state S . See also the example of a design matrix discussed in the Appendix Estimation of GLM.

The simplest domain-based model is to constrain all of the items within the same domain to have identical logit: $h_1(\zeta_{ijs1}) = \xi_{d(j)s1}$, and $h_k(\zeta_{ij sk}, \zeta_{ijs(k-1)}) = \xi_{d(j)sk}, s = 1, \dots, S$. Including the parameter ϕ_{jk} allows deviation of the logit of a specific item from the common logit of the domain. The domain-based model generally has fewer parameters than the model (3.1). The saving in the number of parameters can be substantial when there are many states. Other special cases of (3.3) are possible (e.g., $\phi_{jk} \equiv \phi_j$), but they will not be pursued further here.

(B) Person-specific HMM:

$$h_1(\zeta_{ijs1}) = \phi_{js1} + \theta_i \quad \text{and} \quad h_k(\zeta_{ij sk}, \zeta_{ijs(k-1)}) = \phi_{j sk} + \theta_i, \quad (3.4)$$

where $k = 2, \dots, K_j - 1, \theta_i$ is the individual random effect, and $\theta_j \sim N(0, \sigma^2)$, where σ^2 is the variance of the normal distribution. Model B represents a basic HMM with an additional random effect. The random effect θ_j can be interpreted as a relatively stable trait that consistently affects all of the measured outcomes of an individual.

Clearly, a unifying model can be constructed by combining the domain-based and the person-specific HMMs, and also including person-specific predictors \mathbf{x}_i for the latent trait:

(C) Unified HMM:

$$h_1(\zeta_{ijs1}) = \phi_{j1} + \xi_{d(j)s1} + \theta_i \quad \text{and} \quad (3.5)$$

$$h_k(\zeta_{ijsk}, \zeta_{ijs(k-1)}) = \phi_{jk} + \xi_{d(j)sk} + \theta_i,$$

where $k = 2, \dots, K_j - 1$, and $\theta_i \sim N(\mathbf{x}_i^T \boldsymbol{\gamma}, \sigma^2)$, where $\boldsymbol{\gamma}$ is a p -vector of regression coefficients.

3.1 Parameter Estimation and the M-Step

We first describe the estimating algorithm for the HMM with random effects (Model B). Subsequently, we discuss the estimation of the unified HMM. The EM algorithm continues to be the workhorse for parameter estimation in both cases. Following the notation in the section for the basic HMM estimation, the conditional expected log-likelihood of the complete data $Q(\boldsymbol{\beta}, \boldsymbol{\beta}^{(n)})$ for Model B is given by

$$Q(\boldsymbol{\beta}, \boldsymbol{\beta}^{(n)}) = E[\log p(\mathbf{y}, \mathbf{z}, \boldsymbol{\theta} | \boldsymbol{\beta}) | \mathbf{y}, \boldsymbol{\beta}^{(n)}], \quad (3.6)$$

where $\boldsymbol{\theta} = (\boldsymbol{\theta}_1, \dots, \boldsymbol{\theta}_N)$. Assuming that $\boldsymbol{\theta}$ and \mathbf{z} are independent, we have $p(\mathbf{z}_i, \boldsymbol{\theta}_i | \mathbf{y}_i, \boldsymbol{\beta}^{(n)}) = p(\mathbf{z}_i | \mathbf{y}_i, \boldsymbol{\beta}^{(n)}) f_{\boldsymbol{\theta}}^{(n)}(\boldsymbol{\theta}_i | \mathbf{y}_i)$, where $f_{\boldsymbol{\theta}}^{(n)}(\boldsymbol{\theta}_i | \mathbf{y}_i)$ is the posterior density of $\boldsymbol{\theta}$.

Thus, the objective function is

$$Q(\boldsymbol{\beta}, \boldsymbol{\beta}^{(n)}) = \sum_i \sum_{\mathcal{Z}} \int p(\mathbf{z}_i | \mathbf{y}_i, \boldsymbol{\beta}^{(n)}) \log p(\mathbf{y}_i, \mathbf{z}_i, \boldsymbol{\theta}_i | \boldsymbol{\beta}) f_{\boldsymbol{\theta}}^{(n)}(\boldsymbol{\theta}_i | \mathbf{y}_i) d\boldsymbol{\theta}_i, \quad (3.7)$$

in which the joint distribution of the complete data of subject i within (3.7) is given by

$$p(\mathbf{y}_i, \mathbf{z}_i, \boldsymbol{\theta}_i | \boldsymbol{\beta}) = \alpha_{1z_1} \left(\prod_{t=2}^T \tau_{z_{it-1}z_{it}} \right) \prod_{t=1}^T \prod_{j=1}^J \prod_{k=1}^{K_j} \pi_{ijsk}^{\delta(y_{itj}, k)} f_{\boldsymbol{\beta}}(\boldsymbol{\theta}_i). \quad (3.8)$$

By taking the logarithm on (3.8) and substituting (3.7), we rewrite the objective function:

$$\begin{aligned}
Q(\beta, \beta^{(n)}) &= \sum_i \sum_{\mathcal{Z}} p(\mathbf{z}_i | \mathbf{y}_i, \beta^{(n)}) \log \alpha_{1z_{i1}} \\
&+ \sum_i \sum_{\mathcal{Z}} p(\mathbf{z}_i | \mathbf{y}_i, \beta^{(n)}) \sum_{t=2}^T \log \tau_{z_{it-1}z_{it}} \\
&+ \sum_i \sum_{\mathcal{Z}} p(\mathbf{z}_i | \mathbf{y}_i, \beta^{(n)}) \int \sum_t \sum_j \sum_k \delta(y_{itj}, k) \pi_{ijsk}(\theta_i) f_{\beta^{(n)}}(\theta_i | \mathbf{y}_i) d\theta_i \\
&+ \sum_i \sum_{\mathcal{Z}} p(\mathbf{z}_i | \mathbf{y}_i, \beta^{(n)}) \int \log[f_{\beta}(\theta_i)] f_{\beta^{(n)}}(\theta_i | \mathbf{y}_i) d\theta_i.
\end{aligned} \tag{3.9}$$

In the M-step of the EM algorithm, one needs to evaluate the integrals in (3.9). There is generally no closed-form solution for the integral of the random effects θ_j . Various numerical procedures, including the Gauss–Hermite quadrature (Hinde 1982), Monte Carlo techniques (McCulloch 1997), and Taylor expansions (Wolfinger and O’Connell 1993) have been proposed. Our choice is the Gauss–Hermite quadrature for integration. Denote the vector of q quadrature points by $\mathbf{v} = (v_l)$. The corresponding collection of weights $\mathbf{w}^{(n)} = (w_l^{(n)})$, which depend upon the current estimate of the variance structure, $\sigma^{(n)}$, are used to approximate $f_{\beta^{(n)}}(\theta_j)$. At a quadrature point v_l , denote the conditional probability by $\pi_{ijksl} = [\pi_{ijks}(\theta_j)]_{\theta_j = v_l}$. The integration notwithstanding, computationally the conditional expected log-likelihood still retains the multinomial form. We summarize this observation as a lemma.

Lemma 1—Equation (3.9), which is a generalized form of log-likelihood functions in (2.8) but with an additional dimension of quadrature points for random effects, simplifies to four separate multinomial expressions (2.10). In particular, the first two terms on the right side remain in the same form as the basic HMM. Additionally,

1. in the third term of (3.9), $\mathbf{I} = \{s, j, k, l\}$, $\mathcal{Y}_{\mathbf{I}} = \sum_t w_l^{(n)} \tilde{\alpha}_{itsl}^{(n)} \delta(y_{itj}, k)$, and $\Psi_{\mathbf{I}} = \pi_{ijksl}$ where $\tilde{\alpha}_{itsl}^{(n)} = p(z_{it} = s | \mathbf{y}_i, \beta^{(n)}, v_l)$, of which the conditional density is evaluated at $\theta_j = v_l$; and
2. the fourth term of (3.9) can be simplified into the log-likelihood function of multinomial responses as follows:

$$\sum_i \sum_l w_l^{(n)} \log(w_l), \tag{3.10}$$

where $w_l^{(n)} = f_{\beta^{(n)}}(v_l)$, and $w_l = f_{\beta}(v_l)$.

Proof: Comparing the third row of (3.9) with that of (2.8), we only need to add in $w_l^{(n)}$ and add an extra dimension l to $\tilde{\alpha}^{(n)}$, and thus the proof of the first part is straightforward.

For the second part, because θ_i 's are independent of the latent states, the fourth row of (3.9) simplifies to $\sum_i \int \log[f_\beta(\theta_i)] f_{\beta^{(n)}}(\theta_i) d\theta_i = \sum_i \sum_l \log(w_l) w_l^{(n)}$.

Lemma 1 implies that the computational methods for the basic HMM can still be applied to the random effects model (Model B), but with an additional dimension of quadrature points. For the unified HMM in Model C, there is no closed-form solution for solving the regression coefficients in the conditional probability model. Following the setup in Lemma 1, we can include a generalized linear model (GLM) for the unified HMM. Technical details for estimation of the GLM are provided in an Appendix.

3.2 Computation of the E-Step

The E-step involves the computation of posterior probabilities of the latent states given observed data. Accordingly, the joint distribution and selected marginal distributions have to be evaluated. In HMM, the E-step involves the evaluation of $p(\mathbf{y}_i)$ and $p(\mathbf{z}_i)$. It is well known that brute force summation is not efficient for obtaining marginal probabilities. Various techniques, including approximating algorithms such as Monte Carlo methods, have been proposed (Korb and Nicholson 2003, p. 61). The computations of $p(\mathbf{y}_i)$ and $p(\mathbf{z}_i)$, for example, both require the summation of S^T terms. In our implementation, we adopted the junction-tree method developed from graphical model theory (Jensen, Lauritzen, and Olesen 1990; Lauritzen 1995). The essence of this implementation is a decomposition of a high-dimensional joint probability distribution into self-contained modules of lower dimensions called cliques (to be defined later) and the subsequent local computation of clique probability distribution. In order to do this, the first step is to create a Directed Acyclic Graph (DAG) for the associated statistical model. The directed graph, consisting of nodes (variables) and arcs, is a pictorial representation of the conditional dependence relations of the statistical model. Figure 1(a) shows the DAG of the extended HMM with random effects θ . The second step is to “moralize” the graph, that is, remove arrows and connect (marry) parents, if they are not already connected. A third step is to triangulate the moral graph; edges are added so that chordless cycles contain no more than three nodes. For the basic HMM, the moral and triangulated graphs are topologically identical to the DAG, but for the HMM with random effects, the graphs are not identical. We show the triangulated graph in Figure 1(b). The fourth step is to identify cliques (maximal subsets of nodes that are all interconnected) from the triangulated graph and form compound nodes by fusing nodes of the same clique into a single node. The fifth step is to create separator nodes, which consist of intersections of adjacent compound nodes (Figure 1(c)). The final tree structure is known as a junction tree, an important property of which is that the joint probability function can be factorized into the product of clique marginals over separator marginals (e.g., Jensen, Lauritzen, and Olesen 1990):

$$p(\mathbf{y}) = \frac{\prod_{C \in \mathcal{C}} p(\mathbf{y}_C)}{\prod_{S \in \mathcal{S}} p(\mathbf{y}_S)}, \quad (3.11)$$

where \mathcal{C} and \mathcal{S} , respectively, denote the set of cliques and the set of separators. Cowell et al. (1999, pp. 28–35) provided a simple example to illustrate the process. The junction tree

forms the basis of efficient local computation of clique marginals in the E-step and the global computation of the joint density through so-called message-passing schemes (Pearl 1988; Jensen, Lauritzen, and Olesen 1990), which essentially consists of repeated applications of the Bayes theorem.

Implementation through the junction-tree method has several advantages. First, it has proved to be a highly efficient and scalable algorithm for inference in graphical models, and has been implemented in commercial software such as Hugin (Hugin Expert, Aalborg, Demark). As a special form of the graphical model, the HMM described in this article can be efficiently implemented using junction tree. More importantly, the graphical model framework permits the extension of the basic HMM to include model refinements in highly flexible ways. For example, in the unified model with random effects, nodes representing the stable latent trait can be added to the graphical representation, and the modified junction tree can then be implemented through the aforementioned general algorithm.

The above algorithm was built upon a Matlab toolbox called BNL (Rijmen 2006), which combines the toolbox of Kevin Murphy (Murphy 2001) for constructing the junction tree with a GLM estimation framework for parameterizing the conditional probabilities. The entire program can be downloaded from <http://www.phs.wfubmc.edu/public/downloads/MHMM.zip>.

3.3 Model Selection

In the fitting of HMM models, the number of hidden states is usually not known a priori. An approach suggested by Huang and Bandeen-Roche (2004) in a context of latent class analysis is to employ the Bayesian Information Criterion (BIC; Schwarz 1978) to determine the optimal number of states that gives the lowest BIC value by fitting a range of basic HMMs, without covariates, to the data. While BIC is simple and easy to use, it is a function of the sample size N and has the potential problem of not being able to specify the appropriate “effective sample size” in a longitudinal context because observations of the same subject at different time points are correlated. Generally, if observations from across all T time points are used to determine the number of latent states, the correct sample size should be somewhere between N and NT .

Following the work on the Bayes factor (Kass and Raftery 1995), we derive a criterion that can be used for model selection in the HMM context. Specifically, the criterion is an approximation of the probability density of observations given a specific Model H.

Denote the observed data by \mathbf{Y} and the model parameter by $\boldsymbol{\beta} = (\beta_i) = (\boldsymbol{\alpha}_1, \boldsymbol{\tau}, \boldsymbol{\pi})$, $i = 1, \dots, d$. The marginal probability density of observing the data under Model H can then be written in an integral form:

$$p(\mathbf{Y}|H) = \int p(\mathbf{Y}|\boldsymbol{\beta}, H) f(\boldsymbol{\beta}|H) d\boldsymbol{\beta}, \quad (3.12)$$

where $f(\beta|H)$ is the prior distribution of parameters of Model H. By assuming that the integrand is highly peaked about its maximum at the maximum likelihood estimate $\hat{\beta}$, we can approximate the integral by Laplace's method (De Bruijn 1970), that is,

$$p(\mathbf{Y}|H) \approx (2\pi)^{d/2} |\hat{\Sigma}|^{1/2} p(\mathbf{Y}|\hat{\beta}, H) f(\hat{\beta}|H), \quad (3.13)$$

where $\hat{\Sigma}$ is the covariance matrix, that is, $[-\mathbf{D}^2 \mathcal{K}(\hat{\beta})]^{-1}$, and $\mathcal{K}(\beta) = \log(p(\mathbf{Y}|\beta, H) \times f(\beta|H))$ is the log-likelihood of observed data and the parameter priors.

In HMM, we can explicitly derive from the likelihood functions (2.8) an analytical form of (3.13). Recently, Bayarri et al. (2007) provided a similar derivation for a Bayes factor-based BIC using the multivariate ANOVA model. Visser et al. (2006) derived another Bayes factor under a multivariate autocorrelation model. In the latter work, through simulation experiments, it was also shown that Bayes factor-based criteria perform better than BIC. For the basic HMM (2.8), we follow the notation in Section 2.1 and assume that $K_j \equiv K$. The Laplace approximation leads to the following results.

Lemma 2—*The (-2) times log probability of observing the data, given a discrete hidden Markov model with parameters that follow a Dirichlet prior distribution with a count vector of $\mathbf{1}$, can be approximated with Laplace's method as follows:*

$$-2 \log(p(\mathbf{Y}|H)) \approx -2\tilde{l}(\hat{\beta}) - d \log(2\pi) - \log|\hat{\Sigma}| + 2(S+1) \log(\Gamma(S)) + 2SJ \log(\Gamma(K)),$$

$$(3.14)$$

where $d = S^2 - 1 + SJ(K - 1)$ is the number of parameters of an S -state HMM, $\hat{\Sigma} = [-\mathbf{D}^2 \tilde{\mathcal{K}}(\hat{\beta})]^{-1}$, and $\tilde{l} = \log(p(\mathbf{Y}|\beta, H))$.

Notice that compared to l , \tilde{l} is only the log-likelihood of observed data without the prior parameters. The exact covariance matrix $\hat{\Sigma}$ can be computed using Lystig and Hughes's method (2002) for HMM. A proof is provided in an Appendix. We call the model selection index in (3.14) the modified BIC (mBIC).

To compare the performance of mBIC with BIC, a simulation study was undertaken with the following conditions and levels within each condition: number of states $S = 3, 4$; sample size $N = 100, 200, 400$; number of outcomes $J = 2, 8, 19$; and number of time points $T = 3, 5, 10$. Both 3-state and a 4-state HMMs were used to generate 100 datasets under each condition. The parameters used to generate the data were based on the structure of a 19-item dataset reported by Rijmen et al. (2008). We modified the parameters so that the states are relatively well separated. Table 1 shows the proportion of correct models selected among 100 datasets within each condition. The mBIC tends to perform better than the BIC when any one of the three conditions: sample size N , number of outcomes J , or the number of time points T , is small. For example, for 3-state models, when $N = 100$, $J = 8$, and $T = 3$, mBIC correctly

selected the 3-state model 68% of the time versus 49% for BIC. Both models tended to select the 2-state model when they did not select the correct model. Both the mBIC and BIC perform consistently well when the levels of sample size, number of outcomes, and number of time points become higher.

4. DATA ANALYSIS EXAMPLE

The dataset contains three waves of biennial data collected from 1989 to 1994 in the National Longitudinal Survey of Youth (NLSY79), and the 1990–1994 Children of the NLSY (CoNLSY). Because of missing value problems, which are rather extensive within this NLSY subsample, we only selected those children who had at most two missing values out of a total of 24 measurements taken over the three-year period. This resulted in a sample of 367 children (52% female, 29% African–American, and 20% Hispanic; age range approximately 5–9). There are partially missing values within this final sample, but the percentage is less than 3%. We used multiple imputation (Rubin 1987; Schafer 1997) for handling missing item responses. Before conducting HMM analysis, the dataset was preprocessed—a set of $m = 5$ values were imputed for missing values using PROC MI (SAS Institute Inc.).

Our analysis focuses on cognitive and social development in children five to nine years old. The goal is to identify homogeneous groups of children in terms of their behavioral and cognitive outcomes and to examine how various risk factors, including maternal alcoholism, may affect development. While previous studies (e.g., Snow Jones 2007) have found a positive association between parents' or maternal alcohol consumption and children's behavioral and cognitive problems, many of these studies only assume a linear trend in social or cognitive outcomes as a function of risk factors.

Five outcome variables were selected from the Behavior Problem Index (BPI): anti-social, anxious, peer conflict, headstrong, and hyperactive behaviors. Three variables that measure academic achievement were selected from Peabody Individual Achievement Test (PIAT) scores: math, oral reading, and reading comprehension. We respectively refer to these two sets of variables as representatives of Domains 1 and 2. The BPI and PIAT outcomes were trichotomized into three categories (0, 1, 2) based on nationally normed percentiles with cutpoints at 15% and 85%. Diagnostic groups determined by cutoff points have been commonly used in school reporting systems and research studies to categorize at-risk status. Shepard, Smith, and Vojir (1983), for example, used a cutoff at the 12th percentile in PIAT math to define “below-grade performance” in fifth graders. In a personality study, Hart, Atkins, and Fegley (2003) used the 80th percentiles as cutoff scores on the BPI scales for forming diagnostic groups. We also observe that many school systems use the 15th, 50th, and 85th percentile for reporting purposes, and our choice of cutoffs at 15% and 85% reflects these considerations. To further facilitate model interpretation, BPI categories were then reordered so that movement from 0 to 2 represented increasingly better behavior and conformed to the direction of the PIAT classification.

An important variable that may significantly contribute to a child's development trajectory is home environment. In our analysis, we include home environment variables in the

conditional distributions using the Home Observation for Measurement of the Environment (HOME) inventory. Two HOME scores were available that measured cognitive stimulation and emotional support. Demographic variables such as gender were found not to be significant in this sample, and the analysis results reported here do not include such variables. We report regression models using maternal drinking, and the other HOME variables.

4.1 Results of HMM Analyses

Table 2 shows the BIC and mBIC values for models with two to five hidden states. Both BIC and mBIC values suggest the 5-state model, but both values for the 5-state model are very close to those of the 4-state model. One of the states in the 5-state has low prevalence and was not seen to offer significant information when further consultation with content experts was solicited. We therefore selected the 4-state model for all subsequent analyses. To illustrate the several proposed features of HMM described in Section 2.2, we report results from analyzing the NLSY data using several different HMM models:

Model 0: Basic HMM. No covariate was used, and a saturated conditional probability model was fitted to the data.

Model I: Domain-based HMM. Items that belong to the same domain $d = 1, 2$ (for behavioral and cognitive, respectively) are assumed to follow (3.3).

Model II: HMM with person-specific random effects θ_i to account for dependency given state but without covariate.

Model III: HMM with only fixed effects on conditional probabilities:

$$h_k(\zeta_{isjk}, \zeta_{isj(k-1)}) = \phi_{jk} + \psi_{d(j)sk} + x_{i1}\gamma_1 + x_{i2}\gamma_2 + x_{i3}\gamma_3, \quad (4.1)$$

where x_{i1} , x_{i2} , and x_{i3} are the two HOME scores and the maternal pregnancy drinking score, and β_1 to β_3 are the corresponding fixed effects.

Model IV: Unified HMM with both fixed and random effects on conditional probabilities that follows (3.5).

Table 3 contains the prior probabilities of five states and the transition probabilities table under Model 0, whereas Figure 2 displays the profile of each of the four states, also estimated from Model 0. Each stacked bar in Figure 2 shows the conditional probability distribution of the response categories for a specific item within each state. State 1 can be interpreted as the “no problem or a minor problem” state, State 2 as the “bad behavior (Domain 1) but good academics (Domain 2)” state, State 3 as the “good behavior but poor academics” state, and State 4 as the “bad behavior and poor academics” state. While different models fitted to the data produce different conditional probabilities, the profiles of the four states remain relatively stable (for Models II and IV, the profiles are obtained from averaging over individual profiles). Table 4 reports parameter estimates of Models I–IV and their respective standard errors. The random effects in both Models II and IV are significant. To investigate the heterogeneity of the individuals within a given state, we graphically

display the probabilities of the $N = 367$ individual responses over all the items in Figure 3. The heterogeneity among individuals within a given state appears to be substantial, and the significant variance of θ_j in Models II and IV also reflects the potential conditional dependence among item responses within a given latent state.

From Table 4, the domain-based model suggests that the domain effect—that is, the domain (behavior or academic) to which an item belongs—is highly significant. The domain estimates for the states are consistent across the several models and agree with the findings in Figure 2. For example, the values of domain estimates are higher in State 1 for both academic and behavior, showing that children in State 1 perform better in both domains. State 2 generally has low values in academic and moderately high values in behavior, whereas the values in State 3 are reversed. Models III and IV each contain a regression component for the conditional probabilities using HOME scores and maternal drinking status. Both cognitive stimulation and emotional support at home show strong positive impacts on the conditional response probabilities ($p < 0.001$), indicating that more supportive home environments are associated with better behavior and academic performance. Figure 4 shows that although both high and low Home Cognitive Stimulation groups have increases in State 4 children over time, the percentage of State 1 students is indeed higher in the high Home Cognitive Stimulation group than in the low Home Cognitive Stimulation group. Maternal drinking was not statistically significant. In other words, the long-term stable trait underlying individual behavior and cognition is associated more with home environmental variables and not so much with maternal drinking status. However, it is worth noting that there are only 12 high-risk-pregnancy drinkers in this sample, which may contribute to low power that is not sufficient to detect a difference.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This research was made possible by grants R01 AA11071 (Snow Jones (PI)) from the National Institute of Alcohol Abuse and Alcoholism and the Office of Research on Women's Health, and 0532185 from the National Science Foundation (Ip (PI), Snow Jones (co-PI)).

REFERENCES

- Altman RM. Mixed Hidden Markov Models: An Extension of the Hidden Markov Model to the Longitudinal Data Setting. *Journal of the American Statistical Association*. 2007; 102:201–210.
- Bartolucci F, Pennoni F, Francis B. A Latent Markov Model for Detecting Patterns of Criminal Activity. *Journal of the Royal Statistical Society, Ser. A*. 2007; 170:115–132.
- Baum LE, Petrie T. Statistical Inference for Probabilistic Functions of Finite State Markov Chains. *The Annals of Mathematical Statistics*. 1966; 37:1554–1563.
- Bayarri, S.; Berger, J.; Jang, W.; Pericchi, L.; Ray, S.; Rueda, R.; Visser, I. 6th Workshop on Objective Bayesian Methodology. Roma, Italy: 2007. Extension and Generalization of BIC.
- Chase-Lansdale PL, Mott FL, Brooks-Gunn J, Phillips DA. Children of the National Longitudinal Survey of Youth: A Unique Research Opportunity. *Developmental Psychology*. 1991; 27:918–931.
- Cowell, RG.; Dawid, AP.; Lauritzen, SL.; Spiegelhalter, DJ. *Probabilistic Networks and Expert Systems*. New York: Springer; 1999.

- de Bruijn, NG. *Asymptotic Methods in Analysis*. Amsterdam: North-Holland; 1970.
- Dempster AP, Laird NM, Rubin DB. Maximum Likelihood From Incomplete Data via EM Algorithm. *Journal of the Royal Statistical Society, Ser. B.* 1977; 39:1–38.
- Dias JG, Vermunt JK. Latent Class Modeling of Website Users' Search Patterns: Implications for Online Market Segmentation. *Journal of Retailing and Consumer Services.* 2007; 14:359–368.
- Fahrmeir, L.; Tutz, G. *Multivariate Statistical Modeling Based on Generalized Linear Models*. New York: Springer; 1994.
- Frees, EW. *Longitudinal and Panel Data Analysis and Applications in the Social Sciences*. Cambridge, U.K.: Cambridge University Press; 2004.
- Hart D, Atkins R, Fegley SG. Personality Types of 6-Year-Olds and Their Associations With Academic Achievement and Behavior. *Monographs of the Society for Research in Child Development.* 2003; 68(1):19–34.
- Hinde, JP. Compound Regression Models. In: Gilchrist, R., editor. *GLIM 82: International Conference for Generalized Linear Models*. New York: Springer; 1982. p. 109–121.
- Huang GH, Bandeen-Roche K. Building an Identifiable Latent Variable Model With Covariate Effects on Underlying and Measured Variables. *Psychometrika.* 2004; 69:5–32.
- Hughes JP, Guttorp P, Charles SP. A Nonhomogeneous Hidden Markov Model for Precipitation. *Applied Statistics.* 1999; 48:15–30.
- Humphrey K. The Latent Markov Chain With Multivariate Random Effects: An Evaluation of Instruments Measuring Labor Market Status in the British Household Panel Study. *Sociological Methods and Research.* 1998; 26:269–299.
- Ip EH, Snow Jones A, Heckert DA, Zhang Q, Gondolf E. Latent Markov Model for Analyzing Temporal Configuration for Violence Profiles and Trajectories in a Sample of Batterers. *Sociological Methods and Research.* 2010 to appear.
- Kass RE, Raftery AE. Bayes Factors. *Journal of the American Statistical Association.* 1995; 90(430): 773–795.
- Korb, KB.; Nicholson, AE. *Bayesian Artificial Intelligence*. London: Chapman & Hall; 2003.
- Jensen FV, Lauritzen SL, Olesen KG. Bayesian Updating in Causal Probabilistic Networks by Local Computation. *Computational Statistics Quarterly.* 1990; 4:269–282.
- Lauritzen SL. The EM Algorithm for Graphical Association Models With Missing Data. *Computational Statistics and Data Analysis.* 1995; 19:191–201.
- Laursen B, Hoff E. Person-Centered and Variable-Centered Approaches to Longitudinal Data. *Merrill-Palmer Quarterly.* 2006; 52:377–389.
- Lord, FM.; Novick, MR. *Statistical Theories of Mental Tests Scores*. Readings, MA: Addison-Wesley; 1968.
- Lystig TC, Hughes JP. Exact Computation of the Observed Information Matrix for Hidden Markov Models. *Journal of Computational and Graphical Statistics.* 2002; 11:678–689.
- MacDonald, IL.; Zucchini, W. *Hidden Markov and Other Models for Discrete Valued Time Series*. Boca Raton, FL: Chapman & Hall; 1997.
- McCulloch CE. Maximum Likelihood Algorithms for Generalized Linear Mixed Models. *Journal of the American Statistical Association.* 1997; 92:162–170.
- Murphy, K. *Computing Sciences and Statistics: Proceedings of the 33rd Symposium on the Interface*. Costa Mesa, CA: 2001. The Bayesian Net Toolbox for Matlab.
- Paas LJ, Bijmolt TH, Vermunt JK. Acquisition Patterns of Financial Products: A Longitudinal Investigation. *Journal of Economic Psychology.* 2007; 28:229–241.
- Pearl, J. *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. San Mateo, CA: Kaufman; 1988.
- Rabiner LR. A Tutorial on Hidden Markov-Models and Selected Applications in Speech Recognition. *Proceedings of the IEEE.* 1989; 77:257–286.
- Rijmen, F. *BNL: A Matlab Toolbox for Bayesian Networks With Logistic Regression Nodes*. technical report. 2006. available at <http://www.mathworks.com/matlabcentral/fileexchange/loadFile.do?objectId=13136&objectType=file>

- Rijmen F, Ip E, Rapp S, Shaw E. Qualitative Longitudinal Analysis of Symptoms in Patients With Primary or Metastatic Brain Tumors. *Journal of Royal Statistical Society, Ser. A.* 2008; 171:739–753.
- Rijmen F, Vansteelandt K, De Boeck P. Latent Class Models for Diary Method Data: Parameter Estimation by Local Computations. *Psychometrika.* 2008; 73:163–182.
- Rubin, DB. *Multiple Imputation for Nonresponse in Surveys.* New York: Wiley; 1987.
- Schafer, JL. *Analysis of Incomplete Multivariate Data.* London: Chapman & Hall; 1997.
- Schwarz G. Estimating the Dimension of a Model. *The Annals of Statistics.* 1978; 6:461–464.
- Scott SL, James GM, Sugar CA. Hidden Markov Models for Longitudinal Comparisons. *Journal of the American Statistical Association.* 2005; 100:359–369.
- Shepard LA, Smith ML, Vojir CP. Characteristics of Pupils Identified as Learning Disabled. *American Educational Research Journal.* 1983; 20:309–331.
- Singer, JD.; Willet, JB. *Applied Longitudinal Data Analysis: Modeling Change and Event Occurrence.* New York: Oxford University Press; 2003.
- Snow Jones A. Maternal Alcohol Abuse/Dependence, Children’s Behavior Problems and Home Environment: Estimates From the NLSY Using Propensity Score Matching. *Journal of Studies on Alcohol and Drugs.* 2007; 68(2):266–275. [PubMed: 17286345]
- Snow Jones A, Heckert DA, Gondolf ED, Zhang Q, Ip EH. Complex Behavioral Patterns and Trajectories of Domestic Violence Offenders. *Violence and Victims.* 2010; 25:3–17. [PubMed: 20229690]
- Vermunt JK, Langeheine R, Böckenholt U. Discrete-Time Discrete-State Latent Markov Models With Timeconstant and Time-Varying Covariates. *Journal of Educational and Behavioral Statistics.* 1999; 24:178–205.
- Visser, I.; Ray, S.; Jang, W.; Berger, J. *Workshop on Latent Variable Modeling in Social Sciences.* Research Triangle Park, NC: SAMSI; 2006. Effective Sample Size and the Bayes Factor.
- Wiggins, LM. *Panel Analysis: Latent Probability Models for Attitude and Behavior Processes.* San Francisco, CA: Jossey-Bass; 1973.
- Wolfinger R, O’Connell M. Generalized Linear Mixed Models: A Pseudo-Likelihood Approach. *Journal of Statistical Computation and Simulation.* 1993; 48:233–243.
- Zucchini W, Guttorp P. A Hidden Markov Model for Space Time Precipitation. *Water Resources Research.* 1991; 27:1917–1923.

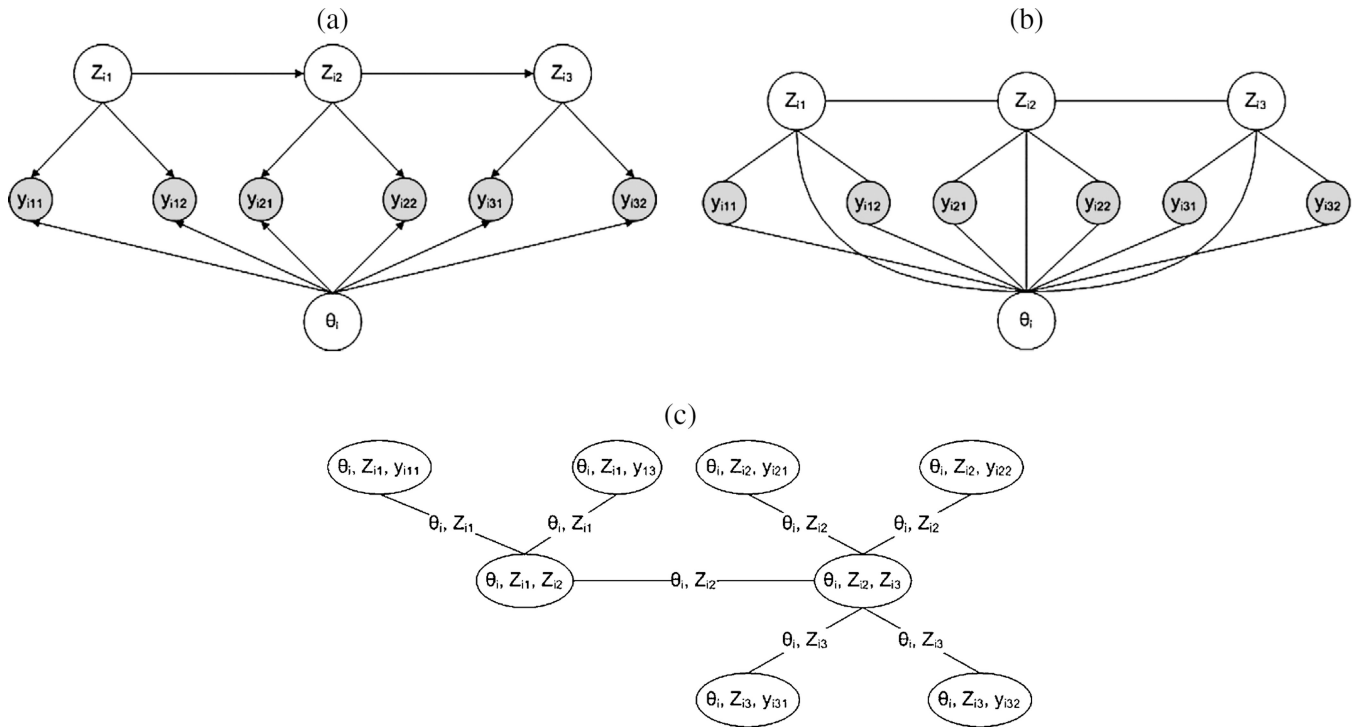


Figure 1. (a) Directed Acyclic Graph (DAG) for HMM with random effects. Shaded ovals indicate observed variables (two are shown here at each time point), and unshaded ovals indicate latent variables. (b) Moral and triangulated graphs for DAG in (a). The parents of the observed variables are joined in a moral graph (e.g., arc from θ to Z_{i1}). (c) Junction tree for DAG in (a). Ovals indicate clique nodes and variables between clique nodes are separators.

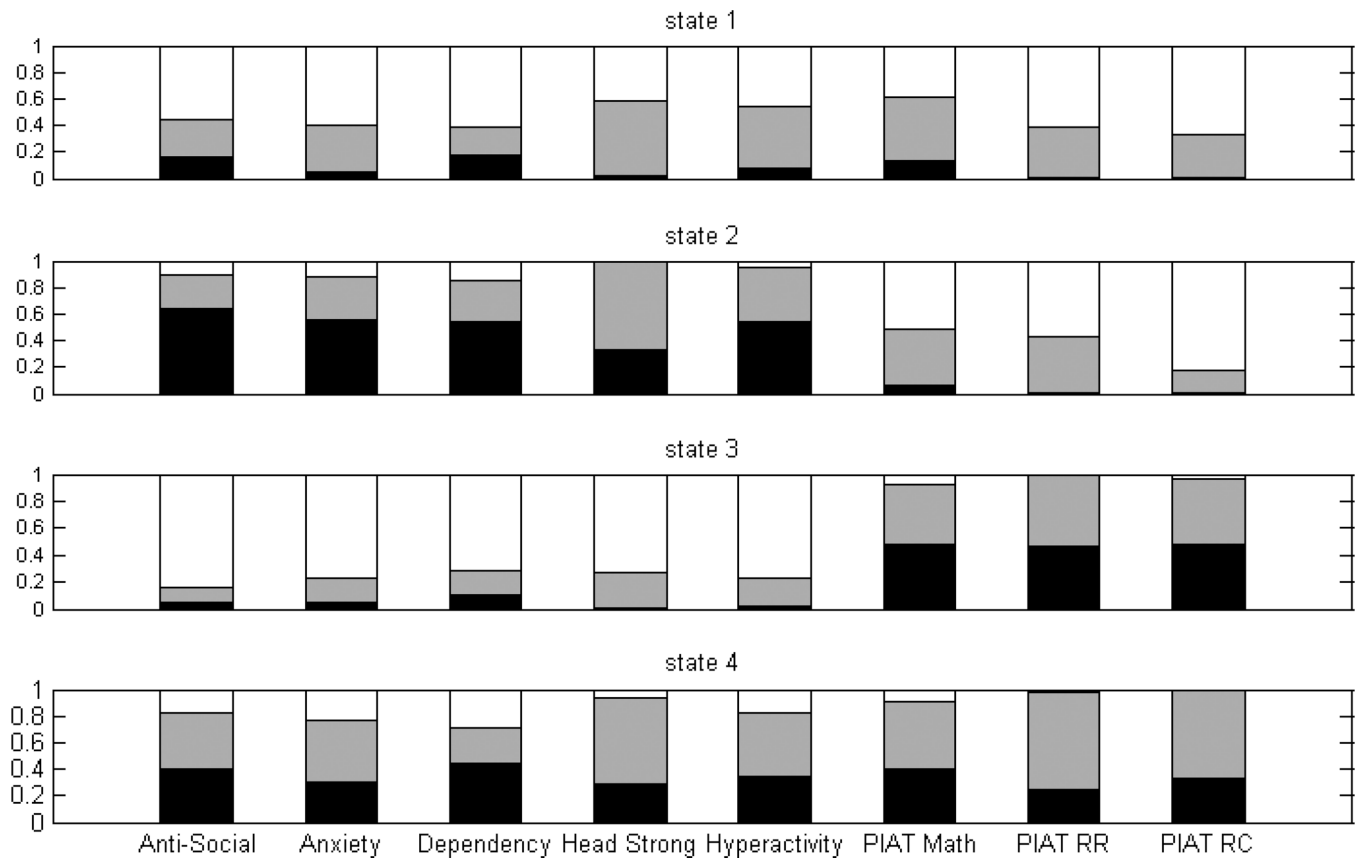


Figure 2. Conditional probability profile for the 4-state model. Each column represents the distribution of expected responses to an item. The bar at the bottom represents the worst response category (0) and the second bar the next worst category (1).

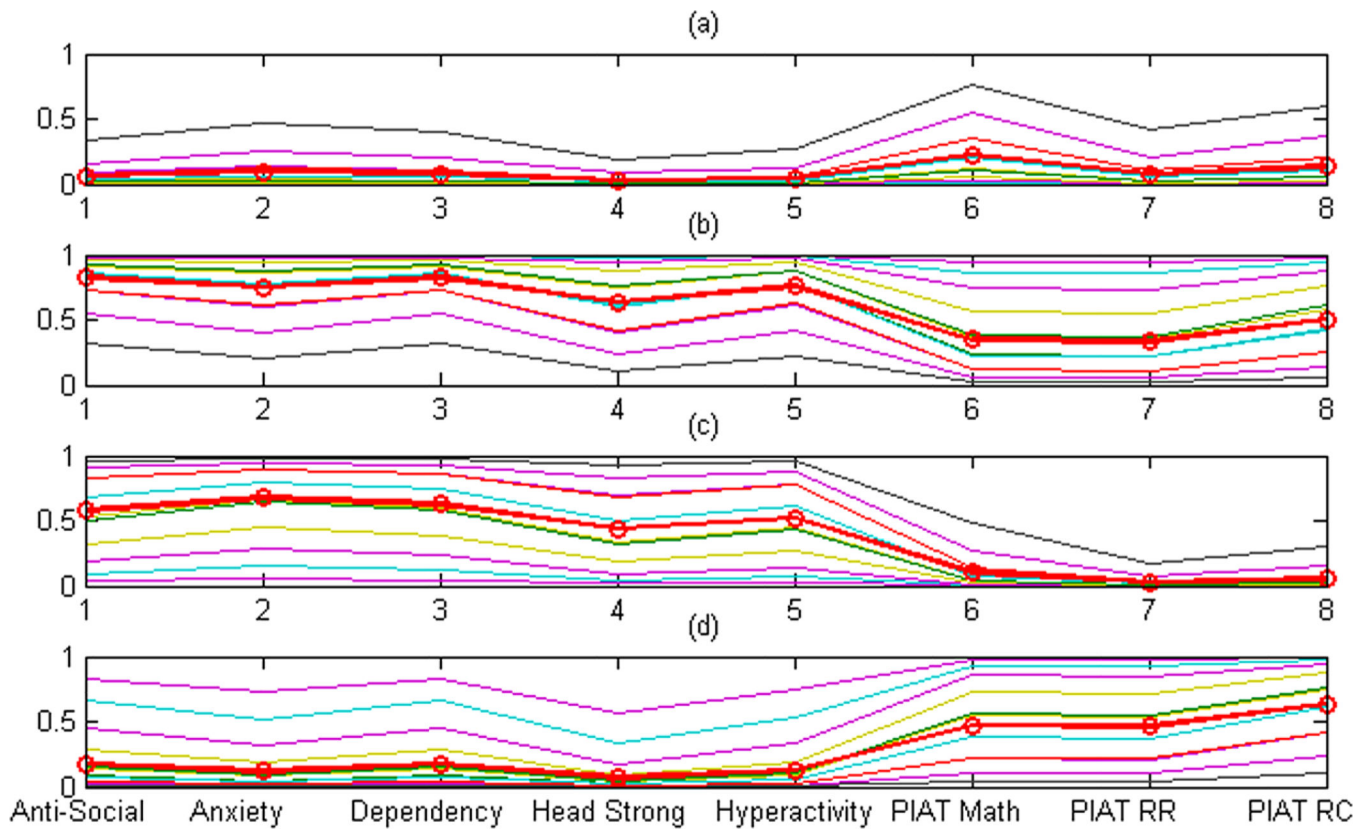


Figure 3. Comparison of sample average conditional probabilities, π_{jsk} , in circles, with fitted individual conditional probabilities, π_{ijsk} , in lines, under Model II. (a) State 1, response category 0; (b) State 1, response category 2; (c) State 2, response category 0; (d) State 2, response category 2. A color version of this figure is available in the electronic version of this article.

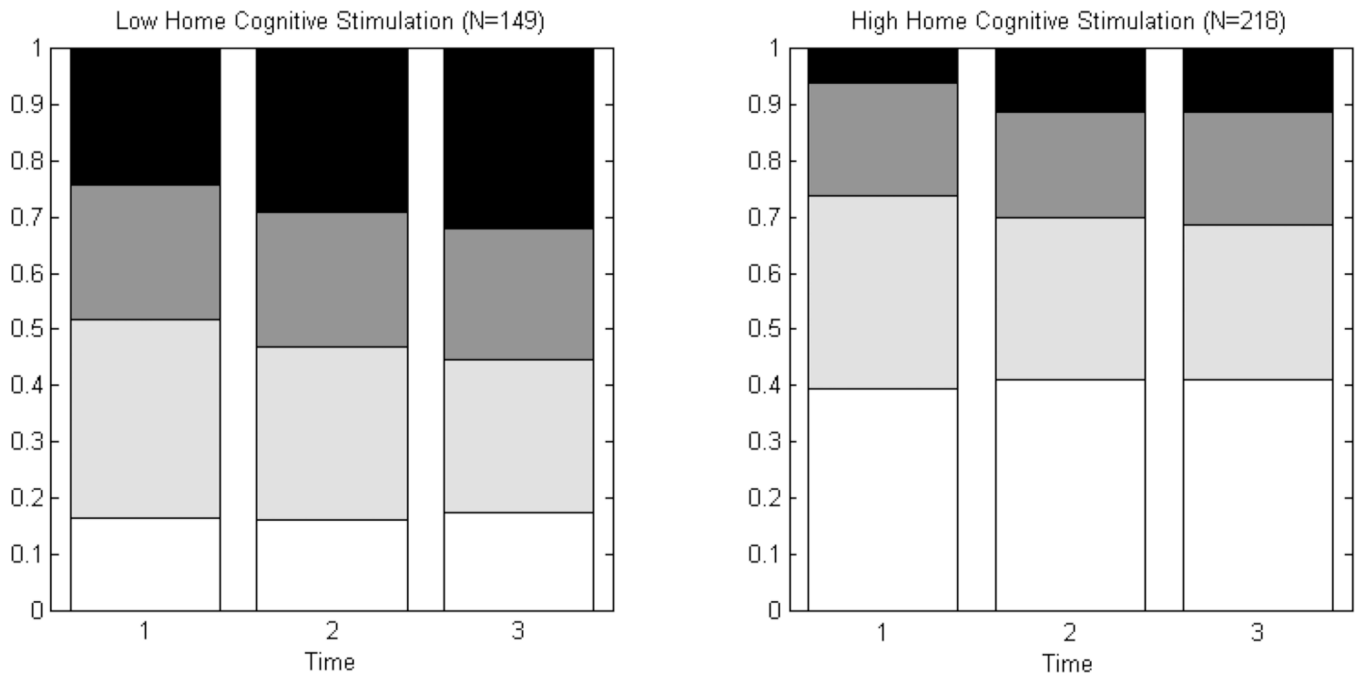


Figure 4. State prevalence over time for two Home Cognitive Stimulation groups. States 1 to 4 are ordered from bottom up.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Model selection simulation results: Proportion of correct models selected for combinations of S states, J outcomes, T time points, with sample size N .

Table 1

S	N	J	T	BIC	mBIC	N	J	T	BIC	mBIC
3	100	2	3	0	0.07	400	2	3	0.01	0.08
	100	2	5	0.23	0.36	400	2	5	1	1
	100	2	10	0.96	0.97	400	2	10	1	0.97
3	100	8	3	0.49	0.68	400	8	3	1	1
	100	8	5	0.9	0.9	400	8	5	1	1
	100	8	10	0.98	0.92	400	8	10	1	0.99
3	100	19	3	1	1	400	19	3	1	1
	100	19	5	0.98	0.98	400	19	5	1	1
	100	19	10	0.92	0.9	400	19	10	1	1
4	200	2	3	0	0.02	400	2	3	0.01	0.12
	200	2	5	0.12	0.11	400	2	5	0.62	0.65
	200	2	10	0.66	0.61	400	2	10	0.87	0.85
4	200	8	3	0.9	0.87	400	8	3	1	1
	200	8	5	1	1	400	8	5	0.99	0.99
	200	8	10	1	1	400	8	10	1	1

Table 2

Number of parameters, deviance, and BIC for the estimated models.

	States	Parameters	Deviance	BIC	mBIC
Basic HMM	2	35	16758	16964	16999
	3	56	16215	16543	16576
	4	79	15819	16284	16324
	5	104	15702	16255	16293
	6	131	15520	16294	16361

Table 3

Estimated prior state probabilities and transition probabilities.

	State 1	State 2	State 3	State 4
Prior probabilities α	0.21	0.31	0.28	0.20
Transition probabilities τ	0.89	0.06	0.03	0.01
	0.07	0.80	0.04	0.09
	0.07	0.11	0.69	0.13
	0.00	0.05	0.13	0.82

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 4

Parameter estimates (standard error) of Model I to Model IV on conditional probabilities. Significant effects at $\alpha = 0.05$ level are highlighted in bold.

	Model I	Model II	Model III	Model IV
Anti-social	Response 1 -1.45 (0.10)	-1.42 (0.17)	-2.08 (0.11)	-1.04 (0.10)
	Response 2 -3.47 (0.21)	-3.80 (0.32)	-3.87 (0.16)	-2.72 (0.11)
Anxiety	Response 1 -1.99 (0.10)	-1.97 (0.17)	-2.60 (0.11)	-1.60 (0.10)
	Response 2 -3.96 (0.21)	-4.36 (0.33)	-4.37 (0.17)	-3.29 (0.12)
Dependency	Response 1 -1.73 (0.10)	-1.70 (0.17)	-2.34 (0.11)	-1.32 (0.10)
	Response 2 -3.45 (0.21)	-3.81 (0.32)	-3.86 (0.16)	-2.73 (0.11)
Head strong	Response 1 -0.72 (0.10)	-0.66 (0.17)	-1.37 (0.11)	-0.26 (0.10)
	Response 2 -4.64 (0.22)	-5.11 (0.33)	-5.07 (0.18)	-4.06 (0.13)
Hyperactivity	Response 1 -1.16 (0.10)	-1.12 (0.17)	-1.79 (0.11)	-0.73 (0.10)
	Response 2 -3.90 (0.21)	-4.29 (0.32)	-4.30 (0.17)	-3.22 (0.12)
Math	Response 1 -0.29 (0.10)	0.25 (0.21)	-1.18 (0.12)	2.34 (0.16)
	Response 2 -3.02 (0.17)	-3.13 (0.28)	-3.87 (0.17)	-0.71 (0.11)
Oral reading	Response 1 1.01 (0.12)	1.79 (0.22)	0.16 (0.13)	3.85 (0.18)
	Response 2 -3.09 (0.17)	-3.19 (0.28)	-3.92 (0.17)	-0.78 (0.11)
Reading comprehension	Response 1 0.37 (0.11)	1.02 (0.21)	-0.51 (0.12)	3.10 (0.17)
	Response 2 -2.15 (0.16)	-2.18 (0.27)	-2.98 (0.15)	0.23 (0.11)
State 1	Domain 1 3.70 (0.12)	3.37 (0.17)	3.51 (0.13)	3.72 (0.25)
	Domain 2 4.00 (0.21)	3.96 (0.32)	3.46 (0.16)	3.70 (0.19)
State 2	Domain 1 3.92 (0.30)	-0.22 (0.22)	3.42 (0.28)	-0.84 (0.31)
	Domain 2 3.54 (0.18)	-0.23 (0.28)	3.19 (0.16)	-0.38 (0.23)
State 3	Domain 1 1.31 (0.09)	1.46 (0.17)	1.05 (0.08)	1.03 (0.08)
	Domain 2 1.65 (0.22)	2.19 (0.32)	1.25 (0.16)	0.87 (0.09)
	Domain 1 3.18 (0.20)	3.28 (0.25)	3.29 (0.20)	-1.73 (0.16)
	Domain 2 2.81 (0.17)	3.55 (0.28)	2.76 (0.15)	-1.93 (0.11)
	Domain 1 2.90 (0.10)	2.56 (0.18)	2.76 (0.10)	2.36 (0.10)
	Domain 2 3.32 (0.21)	3.18 (0.32)	2.91 (0.15)	2.28 (0.10)
	Domain 1 0.50 (0.11)	1.49 (0.22)	0.57 (0.11)	-2.92 (0.17)
	Domain 2 -0.04 (0.21)	1.79 (0.30)	-0.10 (0.19)	-3.00 (0.16)

	Model I	Model II	Model III	Model IV
Home cognitive stimulation			0.83 (0.05)	0.57 (0.05)
Home emotional support			0.12 (0.05)	0.35 (0.05)
Maternal pregnancy drinking			0.52 (0.20)	0.07 (0.22)
Standard deviation of θ_j		0.78 (0.02)		0.70 (0.02)

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript