A TWO-STATE MIXED HIDDEN MARKOV MODEL FOR RISKY TEENAGE DRIVING BEHAVIOR

BY JOHN C. JACKSON[*], PAUL S. ALBERT[†,1] AND ZHIWEI ZHANG[†,1]

United States Military Academy[*] and Eunice Kennedy Shriver National Institute of Child Health and Human Development[†]

This paper proposes a joint model for longitudinal binary and count outcomes. We apply the model to a unique longitudinal study of teen driving where risky driving behavior and the occurrence of crashes or near crashes are measured prospectively over the first 18 months of licensure. Of scientific interest is relating the two processes and predicting crash and near crash outcomes. We propose a two-state mixed hidden Markov model whereby the hidden state characterizes the mean for the joint longitudinal crash/near crash outcomes and elevated g-force events which are a proxy for risky driving. Heterogeneity is introduced in both the conditional model for the count outcomes and the hidden process using a shared random effect. An estimation procedure is presented using the *forward–backward* algorithm along with adaptive Gaussian quadrature to perform numerical integration. The estimation procedure readily yields hidden state probabilities as well as providing for a broad class of predictors.

1. Introduction. The Naturalistic Teenage Driving Study (NTDS), sponsored by the National Institute of Child Health and Human Development (NICHD), was conducted to evaluate the effects of experience on teen driving performance under various driving conditions. It is the first naturalistic study of teenage driving and has given numerous insights to the risky driving behavior of newly licensed teens that includes evidence that risky driving does not decline with experience as discussed by Simons-Morton et al. (2011). During the study 42 newly licensed drivers were followed over the first 18 months after obtaining a license. The participants were paid for their participation, and there were no dropouts. Driving took place primarily in southern Virginia among small cities and rural areas. For each trip, various kinematic measures were captured. A lateral accelerometer recorded driver steering control by measuring g-forces the automobile experiences. These recordings provide two kinematic measures: lateral acceleration and lateral deceleration. A longitudinal accelerometer captured driving behavior along a straight path and records accelerations or decelerations. Another measure for steering control is the

TABLE 1
Kinematic measures and correlation with CNCs, naturalistic teenage driving study, *correlation computed between the CNC and elevated g-force event rates

Category	Gravitation force	Frequency	% Total events	Correlation with CNCs*
Rapid starts	> 0.35	8747	39.6	0.28
Hard stops	≤ -0.45	4228	19.1	0.76
Hard left turns	≤ -0.05	4563	20.6	0.53
Hard right turns	≥ 0.05	3185	14.4	0.62
Yaw	$6°$ in 3 seconds	1367	6.2	0.46
Total		22,090	100	0.60

vehicle's yaw rate, which is the angular deviation of the vehicle's longitudinal axis from the direction of the automobile's path. Each of these kinematic measures was recorded as count data as they crossed specified thresholds that represent normal driving behavior. Crash and near crash outcomes were recorded in two ways. First, the driver of each vehicle had the ability to self report these events. Second, video cameras provided front and rear views from the car during each trip. Trained technicians analyzed each trip the driver took using the video and determination of crash/near crash events made. Table 1 shows the aggregate data for the driving study. More information on the study can be found at http://www.vtti.vt.edu. Our interests are the prediction of crash and near crash events from longitudinal risky driving behavior. Crash or near crash outcomes are our binary outcome of interest, while excessive g-force events are our proxy for risky driving. It is likely that crash/near crash outcomes are best described by some unobserved or latent quality like inherent driving ability. Previously, Jackson et al. (2013) analyzed the driving data using a latent construct where the previously observed kinematic measures describe the hidden state and the hidden state describes the CNC outcome. Our approach here characterizes the joint distribution of crash/near crash and kinematic outcomes using a mixed hidden Markov model where both outcomes contribute to the calculation of the hidden state probabilities.

There is a previous literature on mixed hidden Markov models. Discrete-time mixed Markov latent class models are introduced by Langeheine and van de Pol (1994). A general framework for implementing random effects in the hidden Markov model is discussed by Altman (2007). In this work, the author presented a general framework for a mixed hidden Markov model with a single outcome. The mixed hidden Markov model presented by Altman unifies existing hidden Markov models for multiple processes, which provides several advantages. The modeling of multiple processes simultaneously permits the estimation of population-level effects as well as allowing great flexibility in modeling the correlation structure because they relax the assumption that observations are independent given the hidden states. There are a variety of methods available for estimation of parameters in

mixed hidden Markov models. Altman (2007) performed estimation by evaluating the likelihood as a product of matrices and performing numerical integration via Gaussian quadrature. A quasi-Newton method is used for maximum likelihood estimation. Bartolucci and Pennoni (2007) extend the latent class model for the analysis of capture-recapture data, which takes into account the effect of past capture outcomes on future capture events. Their model allows for heterogeneity among subjects using multiple classes in the latent state. Scott (2002) introduces Bayesian methods for hidden Markov models which was used as a framework to analyze alcoholism treatment trial data [Shirley et al. (2010)]. Bartolucci, Lupparelli and Montanari (2009) use a fixed effects model to evaluate the performance of nursing homes using a hidden Markov model with time-varying covariates in the hidden process. Maruotti (2011) discusses mixed hidden Markov models and their estimation using the expectation–maximization algorithm and leveraging the *forward* and *backward* probabilities given by the *forward–backward* algorithm and partitioning the complete-data-log-likelihood into sub-problems. Maruotti and Rocci (2012) proposed a mixed non-homogeneous hidden Markov model for a categorical data.

We present a model that extends the work of Altman (2007) in several ways. First, we allow the hidden state to jointly model longitudinal binary and count data, which in the context of the driving study represent crash/near crash events and kinematic events, respectively. We also introduce an alternative method to evaluate the likelihood by first using the *forward–backward* algorithm [Baum et al. (1970)] followed by performing integration using adaptive Gaussian quadrature. Implementation of the *forward–backward* algorithm allows for easy recovery of the posterior hidden state probabilities, while the use of adaptive Gaussian quadrature alleviates bias of parameter estimates in the hidden process [Altman (2007)]. Understanding the nature of the hidden state at different time points was of particular interest to us in our application to teenage driving, and our estimation procedure yields an efficient way to evaluate the posterior probability of state occupancy.

In this paper we first introduce a joint model for crash/near crash outcomes and kinematic events which allows the mean of each these to change according to a two-state Markov chain. We introduce heterogeneity in the hidden process as well as the conditional model for the kinematic events via a shared random effect. We then discuss an estimation procedure whereby the likelihood is maximized directly and estimation of the hidden states is readily available by incorporating the *forward–backward* algorithm [Baum et al. (1970)] in evaluating individual likelihoods. We apply our model to the NTDS data and show that these driving kinematic data and CNC events are closely tied via the latent state. We use our model results to form a predictor for future CNC events based on previously observed kinematic data.

2. Methods. Here we present the joint model for longitudinal binary and count outcomes using two hidden states as well as the estimation procedure for model parameters.

2.1. *The model.* Let $\mathbf{b}_i = (b_{i1}, b_{i2}, \ldots, b_{in_i})$ be an unobserved binary random vector whose elements follow a two-state Markov chain (state "0" represents a *good* driving state and state "1" represents a *poor* driving state) with unknown transition probabilities p_{01}, p_{10} and initial probability distribution r_0. We model the crash/near crash outcome Y_{ij}, where i represents an individual and j the month since licensure, using the logistic regression model shown in (1):

(1)
$$Y_{ij} \sim \text{Bernoulli}(\pi_{ij}),$$
$$\text{logit}(\pi_{ij}) = \log(m_{ij}) + \alpha_0 + \alpha_1 b_{ij} + \alpha_2 u_i,$$

where the $\log(m_{ij})$ is an offset to account for the miles driven during a particular month. Treatment of the CNC outcome as binary is not problematic since more than 98% of the monthly observations had one or fewer CNCs observed. Although the log link is ideal for count data, it is a reasonable correction for the miles driven for the logit link when the risk of a crash is low. An alternative parameterization would be to include the miles driven as a covariate in the model. The parameter α_1 assesses the odds-ratio of a crash or near crash event when in the *poor* versus *good* driving state; this odds ratio is simply e^{α_1}, while α_2 reflects unaccounted for covariates beyond the hidden state. The X_{ij} are the sum of the observed elevated g-force count data and this model incorporates heterogeneity among subjects introducing the random effect in the mean structure shown in (2):

(2)
$$X_{ij} \sim \text{Poisson}(\mu_{ij}),$$
$$\log(\mu_{ij}) = \log(m_{ij}) + \beta_0 + \beta_1 b_{ij} + \beta_2 t_j + \beta_3 u_i,$$

where u_i is the random effect with Gaussian distribution and t_j reflects the month of observation since licensure for a particular individual (note that these observations are equally spaced and t_j was not statistically significant when included in the CNC process, hence its omission in this part of the model). Here the parameter β_3 in (2), along with the variance of the random effect distribution, accounts for any variation not explained by the other terms included in the model and induces a correlation between outcomes. Next, we assume that $\{b_{ij}|u_i\}_{j=1}^{n_i}$ is a Markov chain and that $b_{ij}|u_i$ is independent of $b_{it}|u_i$ for $j \neq t$. The transition probabilities for the Markov chain must lie between 0 and 1, and the sum of transitioning from one state to either state must be 1. Hence, the transition probabilities are modeled as

(3)
$$p_{01}(u_i): \quad \text{logit}(\Pr(b_{ij} = 1|b_{i,j-1} = 0|u_i)) = \gamma_{01} + \delta_1 y_{i,j-1} + u_i,$$
$$p_{10}(u_i): \quad \text{logit}(\Pr(b_{ij} = 0|b_{i,j-1} = 1|u_i)) = \gamma_{10} + \delta_2 y_{i,j-1} + \delta^* u_i,$$

where the parameter δ^* in (3) characterizes the degree of correlation between transition probabilities among individuals. Two different types of correlations are described by δ. If $\delta^* < 0$, then this implies individuals have a tendency to remain in either state. If $\delta^* > 0$, then this implies some individuals exhibit a tendency to transition more between states than others. This approach to describe the tran-

sitions in similar to that presented by Albert and Follmann (2007). Introducing random effects in this manner is of great benefit computationally, however, there is a downside risk in that this implementation assumes the processes are highly correlated. This is especially important for the hidden process where if the correlation between transition probabilities is not very strong, biased estimates will result. We present these findings in our simulation section.

2.2. *Estimation.* Letting Ψ represent all parameters included in the model discussed above, the likelihood for the joint model is

$$
L(\Psi; \mathbf{y}, \mathbf{x})
$$

(4)

$$
= \int_{\mathbf{u}} \sum_{\mathbf{b}} f(\mathbf{y}|(\mathbf{b}, \mathbf{u}), \Psi) g(\mathbf{x}|(\mathbf{b}|\mathbf{u}), \Psi) h(\mathbf{u}; \Psi) \, d\mathbf{u}
$$

$$
= \int_{\mathbf{u}} \sum_{\mathbf{b}} \left\{ \prod_{i=1}^{N} \prod_{j=2}^{n_i} f(y_{ij}|(b_{ij}|\mathbf{u}), \Psi) g(x_{ij}|(b_{ij}|\mathbf{u}), \Psi) \right\}
$$

$$
\times \left\{ \prod_{i=1}^{N} r_{i0} \prod_{j=3}^{n_i} \mathrm{p}_{b_{i,j-1}, b_{ij}|\mathbf{u}} \right\} h(\mathbf{u}; \Psi) \, d\mathbf{u},
$$

where the summation associated with $\mathbf{b}$ represents all possible state sequences for an individual and the initial state probabilities are given by $\{r_{i0}\}$ and may include a subject-specific random effect. In (4) we assume the crash/near crash and kinematic event data are conditionally independent. We also assume the $\{u_i\}$ are independent and identically distributed and the observations for any driver are independent given the random effect u_i and the sequence of hidden states. Given these assumptions, the likelihood given in (4) simplifies to a product of one-dimensional integrals shown in (5):

$$
L(\Psi; \mathbf{y}, \mathbf{x})
$$

(5)

$$
= \prod_{i=1}^{N} \int_{u_i} \left\{ \sum_{\mathbf{b}_i} r_{i0} f(y_{i2}|(b_{i2}|u_i), \Psi) g(x_{i2}|(b_{i2}|u_i), \Psi) \right.
$$

$$
\times \prod_{j=3}^{n_i} \mathrm{p}_{b_{i,j-1}, b_{ij}|u_i} f(y_{ij}|(b_{ij}|u_i), \Psi)
$$

$$
\left. \times g(x_{ij}|(b_{ij}|u_i), \Psi) \right\} h(u_i; \Psi) \, du_i.
$$

The different roles the random effects perform in this joint model are of particular interest. The purpose of the inclusion of the random effect in the conditional model for the kinematic data $\{(x_i|(b_i|u_i))\}$ provides a relaxation of the assumption that the observations for an individual are conditionally independent given the

hidden states $\{\mathbf{b}_i\}$ as well as accounting for overdispersion for the kinematic event data. The inclusion of the random effect in the transition probabilities allows the transition probabilities to vary across individuals, inducing a correlation between transition probabilities that induces a relationship between the kinematic events and CNC processes. Further, the random effect provides a departure from the assumption that the transition process follows a Markov chain. One could formulate a reduced model that includes a random effect only in the hidden process, in the hidden process and one or both observed processes, or only in observed processes.

Several possibilities exist for maximizing the likelihood given by (5). Two common approaches are the Monte Carlo expectation maximization algorithm introduced by Wei and Tanner (1990) and the simulated maximum likelihood methods discussed by McCulloch (1997). In Section 2.3, we propose a different method for parameter estimation that does not rely on Monte Carlo methods which are difficult to implement and monitor for convergence. Our method utilizes the *forward–backward* algorithm to evaluate the individual likelihoods conditional on the random effect. As we will show, incorporation of this algorithm provides a simpler means of computing the posterior probability of the hidden state at any time point than MCEM. Further, this method provides a straightforward approach for likelihood and variance evaluation. This approach has the added benefit of producing the estimated variance–covariance matrix for parameter estimates as determined by the inverse of the observed information matrix.

2.3. *Maximizing the likelihood using an implementation of the forward–backward algorithm.* In maximizing the likelihood given by (5), our approach first evaluates the portion of the likelihood described by the observed data given the random effect and hidden states shown here:

$$
(6) \quad \begin{aligned} \Bigg\{ &\sum_{\mathbf{b}_i} r_{i0} f\big(y_{i2}|(b_{i2}|u_i), \Psi\big) g\big(x_{i2}|(b_{i2}|u_i), \Psi\big) \\ &\times \prod_{j=3}^{n_i} \mathrm{p}_{b_{i,j-1},b_{ij}|u_i} f\big(y_{ij}|(b_{ij}|u_i), \Psi\big) g\big(x_{ij}|(b_{ij}|u_i), \Psi\big) \Bigg\}, \end{aligned}
$$

using the *forward–backward* algorithm and subsequent numerical integration over the random effect via adaptive Gaussian quadrature. We then use a quasi-Newton method to maximize this result. The alteration of the *forward–backward* algorithm to accommodate joint outcomes is described as follows. Here, let the vectors $\mathbf{Y}_{ik}^j = (Y_{ik}, \ldots, Y_{ij})'$ with realized values $(y_{ik}, \ldots, y_{ij})'$ and $\mathbf{X}_{ik}^j = (X_{ik}, \ldots, X_{ij})'$ with realized values $(x_{ik}, \ldots, x_{ij})'$. Decompose the joint probability for an individual as follows:

$$
\Pr(b_{ij} = m, \mathbf{Y}_i = \mathbf{y}_i, \mathbf{X}_i = \mathbf{x}_i | u_i)
$$

$$
(7) \quad = \Pr\big(Y_{i2}^j = y_{i2}^j, X_{i2}^j = x_{i2}^j, b_{ij} = m | u_i\big)
$$

$$\times \Pr\big(Y_{i,j+1}^n = y_{i,j+1}^n | Y_{i2}^j = y_{i2}^j, (b_{ij} = m | u_i)\big)$$

$$\times \Pr\big(X_{i,j+1}^n = x_{i,j+1}^n | X_{i2}^j = x_{i2}^j, (b_{ij} = m | u_i)\big)$$

$$= \Pr\big(Y_{i2}^j = y_{i2}^j, X_{i2}^j = x_{i2}^j, b_{ij} = m | u_i\big) \Pr\big(Y_{i,j+1}^n = y_{i,j+1}^n | (b_{ij} = m | u_i)\big)$$

$$\times \Pr\big(X_{i,j+1}^n = x_{i,j+1}^n | (b_{ij} = m | u_i)\big)$$

$$= a_{im}(j) z_{im}(j) \qquad \text{for } m = 0, 1,$$

where $a_{im}(j)$ and $z_{im}(j)$ are referred to as the *forward* and *backward* quantities, respectively, and are

$$a_{im}(j) = \Pr\big(Y_{i2}^j = y_{i2}^j, X_{i2}^j = x_{i2}^j, b_{ij} = m | u_i\big) \qquad \text{for } j = 2, \ldots, n_i,$$

$$a_{im}(1) = \Pr\big((b_{i1} = m | u_i)\big) \Pr\big(Y_{i2} = y_{i2} | (b_{i2} = m | u_i)\big)$$

$$\times \Pr\big(X_{i2} = x_{i2} | (b_{i2} = m | u_i)\big),$$

$$z_{im}(j) = \Pr\big(Y_{i,j+1}^{n_i} = y_{i,j+1}^{n_i}, X_{i,j+1}^{n_i} = x_{i,j+1}^{n_i}, b_{ij} = m | u_i\big)$$

$$\text{for } j = 1, \ldots, (n_i - 1),$$

$$z_{im}(n_i) = 1 \qquad \text{for all } i.$$

The $a_{im}(j)$ and $z_{im}(j)$ are computed recursively in j by using the following:

$$a_{im}(j) = \sum_{l=0}^{1} \Pr\big(Y_{i2}^j = y_{i2}^j, X_{i2}^j = x_{i2}^j, b_{i,j-1} = l, b_{ij} = m, u_i\big)$$

$$= \sum_{l=0}^{1} \Pr\big(Y_{i2}^{j-1} = y_{i2}^{j-1}, X_{i2}^{j-1} = x_{i2}^{j-1}, b_{i,j-1} = l | u_i\big) \mathrm{p}_{lm|u_i}$$

$$\times \Pr\big(Y_{ij} = y_{ij}, X_{ij} = x_{ij} | (b_{ij} = m | u_i)\big)$$

$$= \sum_{l=0}^{1} a_{il}(j-1) \mathrm{p}_{lm|u_i} \Pr\big(Y_{ij} = y_{ij}, X_{ij} = x_{ij} | (b_{ij} = m | u_i)\big)$$

$$= \sum_{l=0}^{1} a_{il}(j-1) \mathrm{p}_{lm|u_i} \Pr\big(Y_{ij} = y_{ij} | (b_{ij} = m | u_i)\big)$$

$$\times \Pr\big(X_{ij} = x_{ij} | (b_{ij} = m | u_i)\big)$$

and

$$z_{im}(j) = \sum_{l=0}^{1} \Pr\big(Y_{i,j+1}^{n_i} = y_{i,j+1}^{n_i}, X_{i,j+1}^{n_i} = x_{i,j+1}^{n_i}, b_{ij+1} = l, (b_{ij} = m | u_i)\big)$$

$$= \sum_{l=0}^{1} \Pr\big(Y_{i,j+1}^{n_i} = y_{i1}^{n_i}, X_{i,j+1}^{n_i} = x_{i,j+1}^{n_i} | (b_{ij+1} = l | u_i)\big) \mathrm{p}_{ml|u_i}$$

$$\times \Pr(Y_{i,j+1} = y_{i,j+1}, X_{i,j+1} = x_{i,j+1} | b_{i,j+1} = l | u_i)$$

$$= \sum_{l=0}^{1} z_{il}(j+1) \mathrm{p}_{ml|u_i} \Pr\big(Y_{i,j+1} = y_{i,j+1}, X_{i,j+1} = x_{i,j+1} | (b_{i,j+1} = l | u_i)\big)$$

$$= \sum_{l=0}^{1} z_{il}(j+1) \mathrm{p}_{ml|u_i} \Pr\big(Y_{i,j+1} = y_{i,j+1} | (b_{i,j+1} = l | u_i)\big)$$

$$\times \Pr\big(X_{i,j+1} = x_{i,j+1} | (b_{i,j+1} = l | u_i)\big).$$

For any individual, the likelihood conditional on the random effect may be expressed as a function of the forward probabilities, so for a two-state Markov chain the conditional likelihood for an individual is

$$L_{i|u_i} = \Pr(\mathbf{Y}_i = \mathbf{y}_i, \mathbf{X}_i = \mathbf{x}_i | u_i)$$

(8)
$$= \sum_{l=0}^{1} \Pr(\mathbf{Y}_i = \mathbf{y}_i, \mathbf{X}_i = \mathbf{x}_i, b_{n_i} = l | u_i)$$

$$= \sum_{l=0}^{1} a_{il}(n_i | \Psi, u_i),$$

where $a_{i0}(n_i | \Psi, u_i)$ and $a_{i1}(n_i | \Psi, u_i)$ are the *forward* probabilities for subject i associated with states 0 and 1, respectively, evaluated at the last observation of the subject's observation sequence n_i. The marginal likelihood for an individual can now be found by integrating with respect to the random effect

(9)
$$L_i = \int_{u_i} \big\{ a_{i0}(n_i | \Psi, u_i) + a_{i1}(n_i | \Psi, u_i) \big\} h(u_i) \, du_i$$

and the complete likelihood can be expressed as a product of individual likelihoods.

2.4. *Numerical integration.* Adaptive Gaussian quadrature can be used to integrate (9). This technique is essential to obtaining accurate parameter estimates, as the integrand is sharply "peaked" at different values depending on the observed measurements of the individual. Applying the results described in Liu and Pierce (1994) to the joint hidden Markov model, numerical integration of (9) is achieved by considering the distribution of the random effects to be $N(0, \theta^2)$. The procedure for obtaining maximum likelihood estimates for model parameters are shown in Table 2.

2.5. *Estimation of posterior hidden state probabilities.* The *forward–backward* implementation in evaluating the likelihood is not only efficient (the number of operations to compute the likelihood conditional on the random effect

TABLE 2
Procedure for obtaining maximum likelihood estimates for the joint mixed hidden Markov model

(1)	Select initial parameter estimates p^0.
(2)	Compute the set of adaptive quadrature points for each individual q_i given the current parameter estimates p^m.
(3)	Maximize the likelihood obtained using the *forward–backward* algorithm and adaptive quadrature via $q_i \in Q$ using the quasi-Newton method.
(4)	Update parameter estimates $p^{(m+1)}$.
(5)	Repeats steps (2)–(4) until parameters converge.

is of linear order as the observation sequence increases), but it also provides a mechanism for recovering information about the hidden states. By leveraging the *forward* and *backward* probabilities, we can compute the hidden posterior state probabilities $\{\hat{b}_{ij}\}$:

$$
\begin{aligned}
\mathrm{E}(b_{ij}|y_i, x_i) &= \mathrm{E}_{u_i|y_i, x_i}\{\mathrm{E}(b_{ij}|y_i, x_i, u_i)\} \\
&= \int_{u_i} \{\Pr(b_{ij} = 1|(y_i, x_i, u_i))\} h_{u_i|(y_i, x_i)}\, du_i \\
&= \int_{u_i} \left\{ \frac{\Pr(b_{ij} = 1, (\mathbf{y}_i, \mathbf{x}_i), u_i)}{\Pr(\mathbf{y}_i, \mathbf{x}_i)} \right\} du_i,
\end{aligned}
$$

(10)

since

$$
\begin{aligned}
\Pr(b_{ij} = 1, (\mathbf{y}_i, \mathbf{x}_i), u_i) &= \Pr(b_{ij} = 1, (\mathbf{y}_i, \mathbf{x}_i)|u_i) h(u_i) \\
&= a_{i1}(j) z_{i1}(j) h(u_i),
\end{aligned}
$$

equation (10) can be expressed as

$$
\begin{aligned}
&\int_{u_i} \left\{ \frac{\Pr(b_{ij} = 1, (\mathbf{y}_i, \mathbf{x}_i), u_i)}{\Pr(\mathbf{y}_i, \mathbf{x}_i)} \right\} du_i \\
&= \int_{u_i} \left\{ \frac{a_{i1}(j) z_{i1}(j) h(u_i)}{\{\int_{u_i} (a_{i0}(j) + a_{i1}(j)) h(u_i)\, du_i\}} \right\} du_i.
\end{aligned}
$$

(11)

Evaluation of (11) is accomplished via adaptive Gaussian quadrature as outlined in the earlier section and the quantities of interest in (11) are readily available after running the *forward–backward* algorithm.

Using a shared random effect is also attractive in that it is computationally more efficient than incorporating separate random effects for the count outcome and transition probabilities. Generally speaking, once the number of random effects exceeds three or four (depending on the type of quadrature being used and the number of nodes included for each integration), direct maximization is no longer a computationally efficient method and Monte Carlo expectation maximization (MCEM) is an appealing alternative approach. In accounting for heterogeneity with a single random effect, we eliminate the need for MCEM.

TABLE 3
Parameter estimates for mixed hidden Markov model 1000 *simulations* (60 *individuals*, 20
observations) *using* $Q = 5$ *and* $Q = 11$ *quadrature points*

Parameter	θ	$Q = 5$			$Q = 11$		
		$\overline{\hat{\theta}}$	$\hat{\theta}_{sd}$	$\sigma_{\hat{\theta}}$	$\overline{\hat{\theta}}$	$\hat{\theta}_{sd}$	$\sigma_{\hat{\theta}}$
α_0	-1.0	-1.01	0.12	0.12	-1.01	0.11	0.11
α_1	2.0	2.03	0.18	0.19	2.03	0.18	0.17
α_2	1.5	1.52	0.41	0.42	1.51	0.42	0.44
β_0	-1.0	-1.01	0.10	0.10	-1.00	0.09	0.06
β_1	2.0	2.01	0.12	0.09	2.02	0.11	0.09
β_2	0.25	0.255	0.06	0.07	0.251	0.06	0.04
γ_{01}	-0.62	-0.61	0.19	0.18	-0.62	0.19	0.17
γ_{10}	0.4	0.42	0.34	0.32	0.42	0.30	0.30
λ	0.0	-0.03	0.16	0.15	-0.03	0.16	0.16
δ^*	2.00	2.15	0.45	0.41	2.02	0.43	0.42
δ_1	1.00	1.08	0.22	0.21	1.02	0.20	0.21
δ_2	3.00	2.97	0.41	0.43	2.97	0.44	0.42
π_{i1}	-0.8	-0.82	0.06	0.05	-0.81	0.05	0.05

3. Simulation of the mixed model. We performed a simulation to investigate the performance of parameter estimation using the proposed approach. Under the shared random effect parameterization, we use the model in (12) for the simulation

$$\text{logit}(\pi_{ij}) = \alpha_0 + \alpha_1 b_{ij} + \alpha_2 u_i,$$

$$\log(\mu_{ij}) = \beta_0 + \beta_1 b_{ij} + \beta_2 u_i,$$

(12) $$\text{logit}\{\Pr(b_{ij} = 1 | b_{ij-1} = 0)\} = \gamma_{01} + \delta_1 y_{i,j-1} + u_i,$$

$$\text{logit}\{\Pr(b_{ij} = 1 | b_{ij-1} = 0)\} = \gamma_{10} + \delta_2 y_{i,j-1} + \delta^* u_i,$$

$$\text{logit}(b_{i1} = 1) = \pi_1,$$

where $u_i \backsim N(0, e^\lambda)$.

The simulations were conducted with 20 observations on 60 subjects. Using a 1.86-GHz Intel Core 2 Duo processor, the fitting of the shared model took less than 3 minutes on average. The simulation results (1000 simulations) are shown in Table 3. Parameter estimates obtained using adaptive Gaussian quadrature with five and eleven points, respectively, are presented along with true parameter value (θ) and mean ($\overline{\hat{\theta}}$), the sample standard deviation for the parameter estimates $\hat{\theta}_{sd}$, and the average asymptotic standard errors $\sigma_{\hat{\theta}}$. In performing the estimation using five quadrature points, parameter estimation was quite accurate with the exception of the coefficient of the random effect δ^* in the 0–1 transition with an average estimated value of 2.15 compared to the actual value of 2.0. Other parameters display very little bias. The bias for δ^* virtually disappears when performing the

integration via adaptive Gaussian quadrature using ten points where the average estimated value was $\delta^* = 2.02$. These results were unchanged when evaluating for possible effects due to the total number of subjects or observations (varying n and I). Additionally, the average asymptotic standard errors agree quite closely to the sample standard deviations for all model parameters. Similar results hold for the case where random effects are only included in the hidden process. We used different starting values to examine the sensitivity of estimation to initial values, and our proposed algorithm was insensitive to the selection of these values. Simulation results provide support that the complexity of the model does not inhibit parameter estimation, rendering discovery of heterogeneity in the observed and hidden processes as a nice byproduct for the model.

While performance of the estimation procedure for the mixed hidden Markov model in (12) was quite good, this model formulation assumes near perfect correlation between the random effects. To examine the robustness of the estimation procedure to this assumption, we evaluated model performance in the case where there are correlated random effects in the hidden process. For this case, data was simulated for the transition probabilities using (13):

$$\text{logit}\{\Pr(b_{ij} = 1 | b_{ij-1} = 0)\} = \gamma_{01} + u_1,$$

(13)

$$\text{logit}\{\Pr(b_{ij} = 1 | b_{ij-1} = 0)\} = \gamma_{10} + u_2,$$

where (u_1, u_2) follow a bivariate normal distribution $BVN(\mathbf{0}, \mathbf{\Sigma})$. Our model was then fit to the simulated data using the parameterization in (14):

$$\text{logit}\{\Pr(b_{ij} = 1 | b_{ij-1} = 0)\} = \gamma_{01} + u_i,$$

(14)

$$\text{logit}\{\Pr(b_{ij} = 1 | b_{ij-1} = 0)\} = \gamma_{10} + \delta u_i.$$

The results for this simulation are shown in Table 4. For moderate departures from perfect correlation, biased estimates result in the hidden process. Thus, we recommend first considering correlated random effects in the hidden process before use of the shared random effects model. In the case of our application, the transition process exhibited very high correlation, so we proceed in the analysis with our estimation procedure. Details for estimation using bivariate adaptive Gaussian quadrature are shown in the supplementary material [Jackson, Albert and Zhang (2015)].

TABLE 4

Parameter estimates for the hidden process when the true underlying random effects distribution is correlated. 1000 simulations (60 individuals, 20 observations)

Parameter	True value	$\rho = 1$	$\rho = 0.95$	$\rho = 0.9$	$\rho = 0.8$	$\rho = 0.7$	$\rho = 0.6$
γ_{10}	-0.62	-0.620	-0.628	-0.640	-0.67	-0.69	-0.72
γ_{01}	0.4	0.401	-0.399	0.398	0.32	0.28	0.26
δ	2.0	2.00	2.11	2.28	2.44	2.61	2.68

TABLE 5
Parameter estimates for the mixed hidden Markov model as applied to the NTDS data

Parameters	Estimate	Std Err
α_0	−7.48	0.14
α_1	1.49	0.25
α_2	0.03	0.02
β_0	−5.97	0.13
β_1	1.31	0.06
β_2	0.007	0.004
β_3	1.10	0.04
λ	−0.18	0.33
δ^*	1.25	0.32
γ_{10}	−2.13	0.35
γ_{01}	−3.47	0.28
δ_1	1.75	0.24
δ_2	−2.17	0.53
π_1	−0.83	0.28

4. Results. The two-state mixed hidden Markov model presented in the previous sections was applied to the NTDS data. Table 5 displays the parameter estimates and associated standard errors. The initial probability distribution for the hidden states was common to all individuals in the study and modeled using $\text{logit}(r_1) = \pi_1$ and the random effects distribution was $N(0, e^\lambda)$. Several models were compared based on the relative goodness-of-fit measure, AIC. Model excursions included evaluating the suitability of a three-state hidden Markov model without random effects ($AIC = 3563.06$) (model shown in the supplementary material [Jackson, Albert and Zhang (2015)]), the two-state model without random effects ($AIC = 3548.97$), and the two-state model with random effects ($AIC = 3492.73$).

The fixed effects model provided initial parameter estimates for most model parameters while multiple starting values for λ were used in conjunction with a grid search over parameters β_3 and δ^* to determine these initial values. The number of quadrature points implemented at each iteration was increased until the likelihood showed no substantial change. As illustrated in the simulation study, there were eleven points used in the adaptive quadrature routine. Standard error estimates were obtained using a numerical approximation to the Hessian using the {NLM} package in R. The coefficients of the hidden states α_1 and β_1 are both significantly greater than zero, indicating that drivers operating in a *poor* driving state ($b_{ij} = 1$) are more likely to have a crash/near crash event and, correspondingly, a highly number of kinematic events. While the variance component of the random effect is somewhat small, the dispersion parameter β_3 is highly significant, indicating the data are overdispersed. Interestingly, heterogeneity is not exhibited

in the CNC outcome as the coefficient for the random effect α_2 is insignificant, providing support to the notion that the hidden state is capturing unobserved quantities in a meaningful way. There is evidence of heterogeneity across individuals in their propensity to change between states as indicated by λ and δ^*. In the case of the NTDS data, $\delta^* > 0$ indicates a positive correlation between the transition probabilities, meaning that some individuals are prone to changing more often between states than others. Coefficients in the hidden process, δ_1 and δ_2, illustrate that transition between states depends on previous CNC outcomes. A prior crash was associated with an increased probability of transitioning from the *good* driving state to the *poor* one ($\delta_1 = 1.75$) and a decreased probability of transitioning from the *poor* to the *good* driving state ($\delta_2 = 2.17$). Since the shared random effect, which assumes a perfect correlation between the random components, may not be robust to a more flexible random effects structure, we also fit the model using correlated random effects in the hidden process (see simulation results). For a variety of starting values, the correlation coefficient estimates were near 1 (0.998 or greater), giving us confidence in using the shared random effect approach.

An interpretation of parameter estimates given in Table 5 is subject-specific and depends on a driver's exposure for a given month. If we consider a subject driving the average mileage for all subjects (358.1 miles), parameter estimates indicate that the risk of a crash/near crash outcome increases from 0.16 to 0.47 when in the *poor* driving state, $b_{ij} = 1$. Correspondingly, this "average" subject would also expect to experience 2.43 more kinematic events on average when in the *poor* driving state. For the typical teenager, the likelihood of moving out of the *poor* driving state decreases from 10.6% to 1.3% when experiencing a CNC event in the previous month. Similarly, the likelihood of moving out of the *good* driving state increases from 3.01% to 15.2% when experiencing a CNC event in the previous month.

A receiver operating characteristic (ROC) curve was constructed to determine the predictive capability of our model by plotting the true positive rate versus the false positive rate for different cutoff values. The ROC curve based on the "one-step ahead" predictions [observed outcome given all previous kinematic observations $\Pr(Y_{ij} = 1 | y_{i1}, \ldots, y_{i,j-1}, x_{i1}, \ldots, x_{i,j-1})$] is shown in Figure 1. An attractive feature of our model is that it allows for the development of a predictor based on prior kinematic events. We constructed this ROC curve using a cross-validation approach whereby one driver was removed from the data set, model parameters were then determined using the remaining data and these results were then used to predict the removed driver's crash/near crash outcomes. The predictive accuracy of this model was moderately high with an area under the curve of 0.74. Although the goodness of fit was best for the two-state mixed hidden Markov model, area under the curve for the other models was nearly identical.

A sample of three drivers and their corresponding hidden state probability $E_{u_i | \mathbf{y}_i, \mathbf{x}_i} \{ E(b_{ij} | \mathbf{y}_i, \mathbf{x}_i, u_i) \}$ along with their crash/near crash outcomes and total number of kinematic events is shown in Figure 2. It is evident how the total kinematic measures influence the predicted value of the hidden state and work particularly well for cases where driving is "consistent" over relatively short time periods
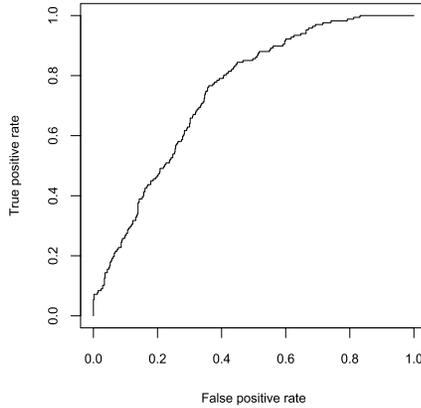
FIG. 1. *ROC curve for the mixed hidden Markov model based one "one-step ahead" predictions (area under the curve = 0.74).*

(i.e., low variation in kinematic measures for a given time period). In cases where the driving kinematics exhibit a great deal of variability, the model does not perform as well in predicting crash/near crash outcomes as indicated by the rightmost panel in Figure 2. As a comparison, we show the results of global decoding of the most likely hidden state sequence using the Viterbi algorithm in Figure 3 for the same three drivers. Hidden state classification is similar whether using global or local decoding for the left two panels in Figure 3, which is indicative of most drivers in the study, while there are differences in the case of the rightmost panel likely due to the greater variability in these data.

5. Discussion. In this paper we presented a mixed hidden Markov model for joint longitudinal binary and count outcomes introducing a shared random effect in the conditional model for the count outcomes and the model for the hidden process. An estimation procedure incorporating the *forward–backward* algorithm with adaptive Gaussian quadrature for numerical integration is used for parameter esti-
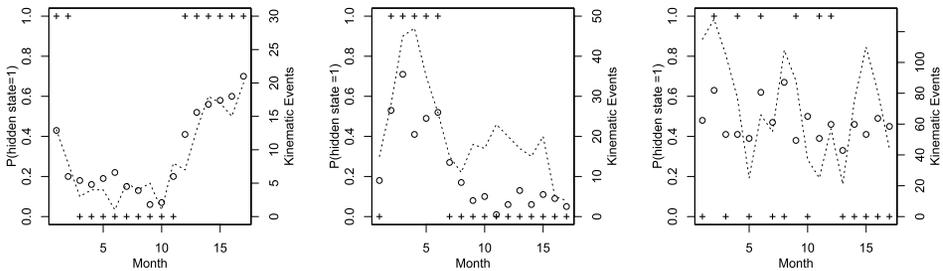
FIG. 2. *Predicted value of the hidden state given the observed data for three drivers. The (○) indicates the probability of being in state 1 (poor driving), (+) indicates a crash/near crash event and the dotted line indicates the composite kinematic measure.*
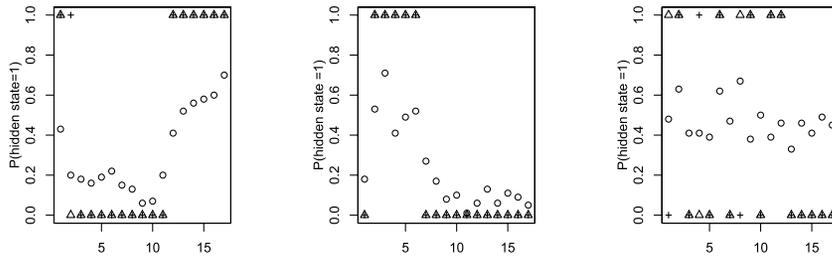
FIG. 3.   *Comparison of local and global decoding of the hidden states and CNC outcomes. The* (○) *indicates the probability of being in state* 1 (*poor driving*), (+) *indicates a CNC event and the* (△) *indicates the hidden state occupation using the Viterbi algorithm.*

mation. A welcome by-product of the *forward–backward* algorithm is the hidden state probabilities for an individual during any time period. The shared random effect eliminates the need for more costly numerical methods in approximating the likelihood, such as higher dimensional Gaussian quadrature or through Monte Carlo EM.

The model was applied to the NTDS data and proved to be a good predictor of crash and near crash outcomes. Our estimation procedure also provides a means of quantifying teenage driving risk. Using the hidden state probabilities which represent the probability of being in a *poor* driving state given the observed crash/near crash and kinematic outcomes, we can analyze the data in a richer way than standard summary statistics. Additionally, our approach allows for a broader class of predictors whereby the investigator may make predictions based on observations that go as far into the past as warranted.

There are limitations to our approach. The shared random effect proposes a rather strong modeling assumption in order to take advantage of an appealing reduction in computational complexity. Using more general correlated random effects approaches is an alternative, but others have found that identification of the correlation parameter is difficult [Smith and Moffatt (1999) and Alfò, Maruotti and Trovato (2011)]. Formal testing for heterogeneity in these models is also a challenging problem [Altman (2008)].

There is also a potential issue of having treated the miles driven during a particular month (m_{ij}) as exogenous. For some crashes, it is possible that previous CNC outcomes ($y_{i,j-1}$) may affect the miles driven in the following month and our model does not capture this dynamic. As with any study, greater clarity in the information obtained for each trip might yield more valuable insights. Metrics such as the type of road, road conditions and trip purpose, while potentially useful, were not available for this analysis.

There are extensions to the model that may be useful. The model can address more than two outcomes. We summarize the kinematic events at a given time as the sum across multiple types. This approach could be extended to incorporate multiple correlated processes corresponding to each kinematic type. Depending on

the situation, the additional flexibility and potential benefits of such an extension may be worth the increased computational cost.

SUPPLEMENTARY MATERIAL

Adaptive quadrature for the three-state mixed hidden Markov model (DOI: 10.1214/14-AOAS765SUPP; .pdf). We provide details on the adaptive quadrature routine for the MHMM with bivariate normal random effects in the hidden process, as well as expressions for the three-state hidden Markov model.

REFERENCES

ALBERT, P. S. and FOLLMANN, D. A. (2007). Random effects and latent processes approaches for analyzing binary longitudinal data with missingness: A comparison of approaches using opiate clinical trial data. *Stat. Methods Med. Res.* **16** 417–439. MR2405479

ALFÒ, M., MARUOTTI, A. and TROVATO, G. (2011). A finite mixture model for multivariate counts under endogenous selectivity. *Stat. Comput.* **21** 185–202. MR2774851

ALTMAN, R. M. (2007). Mixed hidden Markov models: An extension of the hidden Markov model to the longitudinal data setting. *J. Amer. Statist. Assoc.* **102** 201–210. MR2345538

ALTMAN, R. M. (2008). A variance component test for mixed hidden Markov models. *Statist. Probab. Lett.* **78** 1885–1893. MR2528559

BARTOLUCCI, F., LUPPARELLI, M. and MONTANARI, G. E. (2009). Latent Markov model for longitudinal binary data: An application to the performance evaluation of nursing homes. *Ann. Appl. Stat.* **3** 611–636. MR2750675

BARTOLUCCI, F. and PENNONI, F. (2007). A class of latent Markov models for capture-recapture data allowing for time, heterogeneity, and behavior effects. *Biometrics* **63** 568–578. MR2370816

BAUM, L. E., PETRIE, T., SOULES, G. and WEISS, N. (1970). A maximization technique occurring in the statistical analysis of probabilistic functions of Markov chains. *Ann. Math. Statist.* **41** 164–171. MR0287613

JACKSON, J., ALBERT, P. and ZHANG, Z. (2015). Supplement to "A two-state mixed hidden Markov model for risky teenage driving behavior." DOI:10.1214/14-AOAS765SUPP.

JACKSON, J. C., ALBERT, P. S., ZHANG, Z. and SIMONS-MORTON, B. (2013). Ordinal latent variable models and their application in the study of newly licensed teenage drivers. *J. R. Stat. Soc. Ser. C. Appl. Stat.* **62** 435–450. MR3060625

LANGEHEINE, R. and VAN DE POL, F. (1994). Discrete-time mixed Markov latent class models. In *Analyzing Social and Political Change*: *A Casebook of Methods* (A. Dale and R. B. Davies, eds.). Sage, London.

LIU, Q. and PIERCE, D. A. (1994). A note on Gauss–Hermite quadrature. *Biometrika* **81** 624–629. MR1311107

MARUOTTI, A. (2011). Mixed hidden Markov models for longitudinal data: An overview. *International Statistical Review* **79** 427–454.

MARUOTTI, A. and ROCCI, R. (2012). A mixed non-homogeneous hidden Markov model for categorical data, with application to alcohol consumption. *Stat. Med.* **31** 871–886. MR2913866

McCulloch, C. E. (1997). Maximum likelihood algorithms for generalized linear mixed models. *J. Amer. Statist. Assoc*. **92** 162–170. MR1436105

Scott, S. L. (2002). Bayesian methods for hidden Markov models: Recursive computing in the 21st century. *J. Amer. Statist. Assoc*. **97** 337–351. MR1963393

Shirley, K. E., Small, D. S., Lynch, K. G., Maisto, S. A. and Oslin, D. W. (2010). Hidden Markov models for alcoholism treatment trial data. *Ann. Appl. Stat*. **4** 366–395. MR2758176

Simons-Morton, B. G., Ouimet, M. C., Zhang, Z., Klauer, S. E., Lee, S. E., Wang, J., Albert, P. S. and Dingus, T. A. (2011). Crash and risky driving involvement among novice adolescent drivers and their parents. *Am. J. Public Health* **101** 2362–2367.

Smith, M. D. and Moffatt, P. G. (1999). Fisher's information on the correlation coefficient in bivariate logistic models. *Aust. N. Z. J. Stat*. **41** 315–330. MR1718026

Wei, G. C. G. and Tanner, M. A. (1990). A Monte Carlo implementation of the EM algorithm and the poor man's data augmentation algorithms. *J. Amer. Statist. Assoc*. **85** 699–704.

J. C. Jackson
Department of Mathematical Sciences
United States Military Academy
Thayer Hall
West Point, New York 10996
USA
E-mail: john.jackson@usma.edu

P. S. Albert
Z. Zhang
Biostatistics and Bioinformatics
Division of Intramural Population
  Health Research
Eunice Kennedy Shriver National Institute
  of Child Health and Human Development
6100 Executive Blvd
Bethesda, Maryland 20892
USA
E-mail: albertp@mail.nih.gov