

# Web appendix for “A Class of Markov Models for Longitudinal Ordinal Data”

Keunbaik Lee\*

Ph.D. candidate, Department of Statistics, University of Florida, Gainesville FL 32611,  
U.S.A.

and

Michael J. Daniels†

Associate Professor, Department of Statistics, University of Florida, Gainesville FL  
32611, U.S.A.

October 20, 2006

APPENDIX A

*Proof of Theorem I*

To prove Theorem I, we need the following Lemma,

**Lemma.** Under a  $K$  parameter exponential family,  $f(y; \theta) = \exp(\sum_{j=1}^K \theta_j t_j - c(\theta) - d(y))$  where  $\theta = (\theta_1, \dots, \theta_K)$  is the canonical parameter and  $\{t_1(y), \dots, t_K(y)\}$  are the corresponding sufficient statistics,  $\theta_j$  is orthogonal to  $E(t_l)$  for  $j \neq l$ .

---

\* *email:* lee@stat.ufl.edu

† *email:* mdaniels@stat.ufl.edu

*Proof.* Let  $\phi_1 = \theta_j$  and  $\phi_2 = E(t_l) = \frac{\partial c(\theta)}{\partial \theta_l}$ . Then

$$I(\theta_1, \theta_2) = \begin{pmatrix} \frac{\partial^2 c(\theta)}{\partial \theta_j^2} & \frac{\partial^2 c(\theta)}{\partial \theta_j \partial \theta_l} \\ \frac{\partial^2 c(\theta)}{\partial \theta_j \partial \theta_l} & \frac{\partial^2 c(\theta)}{\partial \theta_l^2} \end{pmatrix}. \quad (\text{A.1})$$

We also have

$$I(\theta_1, \theta_2) = \begin{pmatrix} 1 & \frac{\partial^2 c(\theta)}{\partial \theta_j \partial \theta_l} \\ 0 & \frac{\partial^2 c(\theta)}{\partial \theta_l^2} \end{pmatrix} \begin{pmatrix} I_{\phi_1 \phi_1} & I_{\phi_1 \phi_2} \\ I_{\phi_1 \phi_2} & I_{\phi_2 \phi_2} \end{pmatrix} \begin{pmatrix} 1 & 0 \\ \frac{\partial^2 c(\theta)}{\partial \theta_j \partial \theta_l} & \frac{\partial^2 c(\theta)}{\partial \theta_l^2} \end{pmatrix}. \quad (\text{A.2})$$

From (A.1) and (A.2), we have

$$\left( I_{\phi_1 \phi_2} + I_{\phi_2 \phi_2} \frac{\partial^2 c(\theta)}{\partial \theta_j \partial \theta_l} \right) \frac{\partial^2 c(\theta)}{\partial \theta_l^2} = \frac{\partial^2 c(\theta)}{\partial \theta_j \partial \theta_l} \quad (\text{A.3})$$

and

$$I_{\phi_2 \phi_2} \left( \frac{\partial^2 c(\theta)}{\partial \theta_l^2} \right)^2 = \frac{\partial^2 c(\theta)}{\partial \theta_l^2}. \quad (\text{A.4})$$

From (A.4), we have

$$I_{\phi_2 \phi_2} \frac{\partial^2 c(\theta)}{\partial \theta_l^2} = 1. \quad (\text{A.5})$$

From (A.3) and (A.5), we have

$$I_{\phi_1 \phi_2} \frac{\partial^2 c(\theta)}{\partial \theta_l^2} + \frac{\partial^2 c(\theta)}{\partial \theta_j \partial \theta_l} = \frac{\partial^2 c(\theta)}{\partial \theta_j \partial \theta_l}.$$

Since  $\frac{\partial^2 c(\theta)}{\partial \theta_l^2} > 0$ ,  $I_{\phi_1 \phi_2} = 0$ . So  $\theta_j$  is orthogonal to  $E(t_l)$ .

*Proof of Theorem I:* Using the fact that the likelihood contribution from each subject can be factored into the product of the probability for the first observation times the subsequent, one-step transition probabilities allows a sequential calculation of the likelihood. Consider a vector of responses from a subject  $i$ ,  $Y_i = (Y_{i1}, \dots, Y_{in})$ . To obtain

the likelihood function for subject  $i$ , we compute the sequential products of transition probabilities,

$$\begin{aligned}
& P(Y_{i1} = y_{i1}) \\
&= \prod_{k=1}^K (P_{i1k}^M - P_{i1k-1}^M)^{y_{i1k}} \\
&= \exp \left[ \{y_{i11}\phi_{i11} - (y_{i11} + y_{i12})g(\phi_{i11})\} + \cdots + \left\{ \sum_{k=1}^{K-1} y_{i1k}\phi_{i1K-1} - \sum_{k=1}^K y_{i1k}g(\phi_{i1K-1}) \right\} \right] \\
&= \exp \left\{ -g(\phi_{i1K-1}) + \sum_{g=1}^{K-1} \left( \sum_{k=g}^{K-1} \phi_{i1k} - \sum_{k=1}^{K-2} g(\phi_{i1k}) \right) y_{i1g} \right\} \\
&= \exp \left( \theta_{i01} + \sum_{g=1}^{K-1} \eta_{i1g} y_{i1g} \right)
\end{aligned}$$

where  $\phi_{i1k} = \log \left( \frac{P_{i1k}^M}{P_{i1k+1}^M - P_{i1k}^M} \right)$ ,  $g(a) = \log(1 + e^a)$ ,  $\eta_{i1g} = \sum_{k=g}^{K-1} \phi_{i1k} - \sum_{k=1}^{K-2} g(\phi_{i1k})$ , and  $\theta_{i01} = -g(\phi_{i1K-1})$ . Then

$$\begin{aligned}
& P(Y_{i1} = y_{i1}, Y_{i2} = y_{i2}) = P(Y_{i1} = y_{i1})P(Y_{i2} = y_{i2}|Y_{i1} = y_{i1}) \\
&= \exp \left( \theta_{i01} + \sum_{g=1}^{K-1} \eta_{i1g} y_{i1g} \right) \exp \left\{ -\log \left( 1 + \sum_{m=1}^{K-1} e^{\Delta_{i2m} + \gamma_{i211}^{(m)} y_{i11} + \cdots + \gamma_{i21K-1}^{(m)} y_{i1K-1}} \right) \right. \\
&\quad \left. + \sum_{k=1}^{K-1} \left( \Delta_{i2k} + \gamma_{i211}^{(k)} y_{i11} + \cdots + \gamma_{i21K-1}^{(k)} y_{i1K-1} \right) y_{i2k} \right\} \\
&= \exp \left\{ \theta_{i02} + \sum_{k=1}^{K-1} \eta_{i1k} y_{i1k} + \sum_{k=1}^{K-1} \Delta_{i2k} y_{i2k} + \sum_{k=1}^{K-1} \left( \gamma_{i211}^{(k)} y_{i11} y_{i2k} + \cdots + \gamma_{i21K-1}^{(k)} y_{i1K-1} y_{i2k} \right) \right. \\
&\quad \left. + \log \left( 1 + \sum_{m=1}^{K-1} e^{\Delta_{i2m}} \right) - \log \left( 1 + \sum_{m=1}^{K-1} e^{\Delta_{i2m} + \gamma_{i211}^{(m)} y_{i11} + \cdots + \gamma_{i21K-1}^{(m)} y_{i1K-1}} \right) \right\} \\
&= \exp \left\{ \theta_{i02} + \sum_{k=1}^{K-1} \theta_{i1k} y_{i1k} + \sum_{k=1}^{K-1} \Delta_{i2k} y_{i2k} + \sum_{k=1}^{K-1} \left( \gamma_{i211}^{(k)} y_{i11} y_{i2k} + \cdots + \gamma_{i21K-1}^{(k)} y_{i1K-1} y_{i2k} \right) \right\},
\end{aligned}$$

where  $\theta_{i02} = \theta_{i01} - \log\left(1 + \sum_{m=1}^{K-1} e^{\Delta_{i2m}}\right)$  and  $\theta_{i1k} = \eta_{i1k} + \log\left(1 + \sum_{m=1}^{K-1} e^{\Delta_{i2m}}\right) - \log\left(1 + \sum_{m=1}^{K-1} e^{\Delta_{i2m} + \gamma_{i21k}^{(m)}}\right)$ , for  $k = 1, \dots, K-1$ . Finally, we have the following equation

$$\begin{aligned} & P(Y_{i1} = y_{i1}, \dots, Y_{iT} = y_{iT}) \\ &= \exp \left\{ \theta_{i0T} + \sum_{j=1}^T \sum_{k=1}^{K-1} \theta_{ijk} y_{ijk} + \sum_{j=2}^{n_i} \sum_{k=1}^{K-1} (\gamma_{ij11}^{(k)} y_{ij-11} y_{ijk} + \dots + \gamma_{ij1K-1}^{(k)} y_{ij-1K-1} y_{ijk}) \right\} \end{aligned} \quad (\text{A.6})$$

where

$$\begin{aligned} \theta_{i0T} &= -\log\left(1 + \sum_{m=1}^{K-1} e^{\eta_{i1k}}\right) - \sum_{j=2}^T \log\left(1 + \sum_{m=1}^{K-1} e^{\Delta_{ijm}}\right), \\ \theta_{ijk} &= \Delta_{ijk} + \log\left(1 + \sum_{m=1}^{K-1} e^{\Delta_{ij+1m}}\right) - \log\left(1 + \sum_{m=1}^{K-1} e^{\Delta_{ij+1m} + \gamma_{ij+11k}^{(m)}}\right), \\ & \text{and} \\ \theta_{iT k} &= \Delta_{iT k}. \end{aligned}$$

Since (A.6) is a canonical exponential form,  $\gamma(\alpha)$  is orthogonal to  $E(y_{ijk})$  by Lemma 1. Note that  $E(y_{ijk})$  is a function of only  $\beta_0$  and  $\beta$ . By Barndorff-Nielsen and Cox (1994, p 50),  $\alpha$  is orthogonal to  $\beta_0$  and  $\beta$ .

## APPENDIX B

### *Proof of Corollary I*

Let  $c(y_{it-1}; \alpha^{(k)}, q) = \gamma_{it11}^{(k)} y_{it-1} + \gamma_{it12}^{(k)} y_{it-1}^2 + \dots + \gamma_{it1s}^{(k)} y_{it-1}^s$ . Replace  $c(y_{it-1}; \alpha^{(k)}, s)$  with  $\gamma_{itj}^{*(k)}$ . Then, we have

$$\log\left(\frac{P_{itk}^c}{P_{itK}^c}\right) = \Delta_{itk} + \gamma_{it1j}^{*(k)} y_{it-11} + \dots + \gamma_{it1j}^{*(k)} y_{it-1K}.$$

Re-express the above equation as

$$\log\left(\frac{P_{itk}^c}{P_{itK}^c}\right) = \Delta_{itk} + \gamma_{it1j}^{*(k)} y_{it-11}^* + \dots + \gamma_{it1j}^{*(k)} y_{it-1K-1}^*,$$

where  $y_{it-1j}^* = (y_{it-1j} + \frac{1}{K-1}y_{it-1K})$ . Then proceed as in the proof of Theorem I.

## APPENDIX C

### *Proof of Theorem II*

From the proof of Theorem I, we obtain the joint distribution and re-express it as

$$P(y_i; \theta, \gamma) = \exp \{ \theta_i^T y_i + \gamma_i^T w_i + \theta_{i0} \}, \quad (\text{C.7})$$

where

$$\begin{aligned} y_i^T &= (y_{i1}^T, \dots, y_{iT}^T); \quad y_{it}^T = (y_{it1}, \dots, y_{itK-1}), \\ w_i^T &= (w_{i2}^T, \dots, w_{iT}^T, \omega_{i123}^T, \dots, \omega_{i1\dots T}^T); \quad w_{it}^T = (w_{it1}^T, \dots, w_{itK-1}^T); \\ w_{itk}^T &= (y_{it-11}y_{itk}, \dots, y_{it-1K-1}y_{itk}), \\ \omega_{i123}^T &= (y_{i11}y_{i21}y_{i31}, \dots, y_{iT-2K-1}y_{iT-1K-1}y_{iTK-1}), \dots, \\ \omega_{i1\dots T}^T &= (y_{i11}y_{i21} \dots y_{iT1}, \dots, y_{iT-2K-1}y_{iT-1K-1} \dots y_{iTK-1}), \\ \theta_i^T &= (\theta_{i1}^T, \dots, \theta_{iT}^T); \quad \theta_{it}^T = (\theta_{it1}, \dots, \theta_{itK-1}), \\ \gamma_i^T &= (\gamma_{i2}^T, \dots, \gamma_{iT}^T, 0^T, \dots, 0^T); \quad \gamma_{it}^T = (\gamma_{it}^{(1)T}, \dots, \gamma_{it}^{(K-1)T}); \\ \gamma_{it}^{(k)T} &= (\gamma_{it11}^{(k)}, \dots, \gamma_{it1K-1}^{(k)}), \\ \theta_{i0} &= -\log \left( 1 + \sum_{m=1}^{K-1} e^{\eta_{i1k}} \right) - \sum_{j=2}^T \log \left( 1 + \sum_{m=1}^{K-1} e^{\Delta_{ijm}} \right). \end{aligned}$$

The log likelihood for the  $i$ th individual is

$$l_i = \log P(y_i; \theta, \gamma) = \theta_i^T y_i + \gamma_i^T w_i + \theta_{i0}.$$

Let  $\beta^{*T} = (\beta_0^T, \beta^T)$ . Then the likelihood equations can be written,

$$\begin{pmatrix} \frac{\partial l_i}{\partial \beta^*} \\ \frac{\partial l_i}{\partial \alpha} \end{pmatrix} = \begin{pmatrix} \frac{\partial P_i^M}{\partial \beta^*} & \frac{\partial P_i^M}{\partial \alpha} \\ \frac{\partial \nu_i}{\partial \beta^*} & \frac{\partial \nu_i}{\partial \alpha} \end{pmatrix}^T \begin{pmatrix} V_{i11} & V_{i12} \\ V_{i21} & V_{i22} \end{pmatrix}^{-1} \begin{pmatrix} y_i - P_i^M \\ w_i - \nu_i \end{pmatrix},$$

where

$$\begin{pmatrix} V_{i11} & V_{i12} \\ V_{i21} & V_{i22} \end{pmatrix} = \begin{pmatrix} \text{cov}(Y_i) & \text{cov}(Y_i, W_i) \\ \text{cov}(W_i, Y_i) & \text{cov}(W_i) \end{pmatrix}, \quad P_i^M = E(y_i), \quad \nu_i = E(W_i).$$

Thus, we have score functions as follows

$$\begin{aligned} \sum_{i=1}^N X_i^{*T} A_i V_{i11}^{-1} (y_i - P_i^M) &= 0 \\ \sum_{i=1}^N Z_i^T \{ (w_i - \nu_i) - V_{i21} V_{i11}^{-1} (y_i - P_i^M) \} &= 0, \end{aligned} \tag{C.8}$$

where  $X_i^*$  is a  $(K-1)T \times (K-1+r)$  matrix such as

$$X_i^* = \begin{pmatrix} 1 & 0 & \cdots & 0 & x_{i1}^T \\ \vdots & \cdots & & & \vdots \\ 0 & 0 & \cdots & 1 & x_{i1}^T \\ \vdots & \cdots & & & \vdots \\ 1 & 0 & \cdots & 0 & x_{iT}^T \\ \vdots & \cdots & & & \vdots \\ 0 & 0 & \cdots & 1 & x_{iT}^T \end{pmatrix},$$

$Z_i$  is a  $((T-1)(K-1)^2q + (K-1)^3 + \cdots + (K-1)^T) \times ((T-1)(K-1)^2q + (K-1)^3 + \cdots + (K-1)^T)$  matrix such as

$$Z_i^T = \begin{pmatrix} z_{i2}^T, \cdots, z_{i2}^T, \cdots, z_{iT}^T, \cdots, z_{iT}^T, 0^T, \cdots, 0^T \\ \vdots \\ z_{i2}^T, \cdots, z_{i2}^T, \cdots, z_{iT}^T, \cdots, z_{iT}^T, 0^T, \cdots, 0^T \\ 0^T, \cdots, 0^T, \cdots, 0^T, \cdots, 0^T, 0, \cdots, 0^T \\ \vdots \\ 0^T, \cdots, 0^T, \cdots, 0^T, \cdots, 0^T, 0, \cdots, 0^T \end{pmatrix}$$

and

$$A_i = \text{diag} \left\{ \left( \frac{1}{P_{i11}^M} - \frac{1}{P_{i1K}^M} \right)^{-1}, \cdots, \left( \frac{1}{P_{iT_{K-1}}^M} - \frac{1}{P_{iT_K}^M} \right)^{-1} \right\}.$$

Taking the derivative of (C.8) with respect to  $\alpha$  and expectation with respect to the true  $\beta^*$  and arbitrary  $\alpha_*$ , we have

$$\sum_{i=1}^N \frac{\partial}{\partial \alpha} (X_i^* A_i V_{i11}^{-1}) (E_{\beta^*, \alpha_*}(y_i) - P_i^M) = 0.$$

Thus, we have an unbiased estimating equation for  $\beta^*$  for all  $\alpha$  (Firth, 1987). This implies that the MLE for  $\beta$  is a consistent estimator regardless of the value  $\alpha$ . Note that OMTM(1) and IOMTM(1) are special cases of (C.7).

## APPENDIX D

### *Detailed calculations for Fisher-scoring for OMTM(2)*

The contribution of  $L^{(1)}$  and  $L^{(2)}$  to the score equations is almost identical to that described for OMTM(1) in the Supplementary Materials. The contribution of  $L^{(3)}$  is given by

$$\begin{aligned} \frac{\partial \log L^{(3)}}{\partial \theta_1} &= \sum_{i=1}^N \sum_{t=3}^{n_i} \sum_{k=1}^{K-1} (y_{itk} - P_{itk}^c) \frac{\partial \Delta_{itk}}{\partial \theta_1}, \\ \frac{\partial \log L^{(3)}}{\partial \alpha_{lb}^{(d)}} &= \sum_{i=1}^N \sum_{t=3}^{n_i} \left\{ \sum_{k=1}^{K-1} (y_{itk} - P_{itk}^c) \frac{\partial \Delta_{itk}}{\partial \alpha_{lb}^{(d)}} + (y_{ita} - P_{ita}^c) y_{it-lb} z_{it} \right\}, \end{aligned}$$

for  $b = 1, \dots, K-1$ ,  $d = 1, \dots, K-1$ , and  $l = 1, \dots, K-1$ .

Now for the information matrix, we have the following results,

$$\begin{aligned}
E_y \left( -\frac{\partial^2 \log L^{(3)}}{\partial \theta_1 \partial \theta_1^T} \right) &= E_y \left( \sum_{i=1}^N \sum_{t=3}^{n_i} \sum_{k=1}^{K-1} \sum_{g=1}^{K-1} \frac{\partial P_{itk}^c}{\partial \Delta_{itg}} \frac{\partial \Delta_{itk}}{\partial \theta_1} \frac{\partial \Delta_{itg}}{\partial \theta_1^*} \right), \\
E_y \left( -\frac{\partial^2 \log L^{(3)}}{\partial \alpha_{ab}^{(d)} \partial \alpha_{ce}^{(f)T}} \right) &= E_y \left[ \sum_{i=1}^N \sum_{t=3}^{n_i} \left\{ \sum_{k=1}^{K-1} \left( \sum_{g=1}^{K-1} \frac{\partial P_{itk}^c}{\partial \Delta_{itg}} \frac{\partial \Delta_{itg}}{\partial \alpha_{ab}^{(d)}} \frac{\partial \Delta_{itk}}{\partial \alpha_{ce}^{(f)T}} + \frac{\partial P_{itk}^c}{\partial \gamma_{itab}^{(d)}} \frac{\partial \gamma_{itab}^{(d)}}{\partial \alpha_{ab}^{(d)}} \frac{\partial \Delta_{itk}}{\partial \alpha_{ce}^{(f)T}} \right) \right. \right. \\
&\quad \left. \left. + \sum_{g=1}^{K-1} \frac{\partial P_{itf}^c}{\partial \Delta_{itg}} \frac{\partial \Delta_{itg}}{\partial \alpha_{ab}^{(d)}} y_{it-1b} z_{it}^T + \frac{\partial P_{itf}^c}{\partial \gamma_{itab}^{(d)}} \frac{\partial \gamma_{itab}^{(d)}}{\partial \alpha_{ab}^{(d)}} y_{it-1f} z_{it}^T \right\} \right], \\
E_y \left( -\frac{\partial^2 \log L^{(3)}}{\partial \alpha_{ab}^{(d)} \partial \theta_1} \right) &= E_y \left[ \sum_{i=1}^N \sum_{t=3}^{n_i} \sum_{k=1}^{K-1} \left\{ \sum_{g=1}^{K-1} \frac{\partial P_{itk}^c}{\partial \Delta_{itg}} \frac{\partial \Delta_{itg}}{\partial \alpha_{ab}^{(d)}} \frac{\partial \Delta_{itk}}{\partial \theta_1} + \frac{\partial P_{itk}^c}{\partial \gamma_{itab}^{(d)}} \frac{\partial \gamma_{itab}^{(d)}}{\partial \alpha_{ab}^{(d)}} \frac{\partial \Delta_{itk}}{\partial \theta_1} \right\} \right], \tag{D.9}
\end{aligned}$$

for  $a, c = 1, \dots, p$  and  $b, d, e, f = 1, \dots, K-1$ .

We also need derivatives of  $\Delta_{it}$  with respect to  $\beta_0$ ,  $\beta$ , and  $\alpha$ . They can be obtained as the solution to the following system of linear equations,

$$\begin{aligned}
\frac{\partial \Delta_{it1}}{\partial \theta_1} \sum_{g_1=1}^K \sum_{g_2=1}^K \frac{\partial h_{ikg_1g_2}^{(t)}}{\partial \Delta_{it1}} P_{it-1g_1g_2}^J + \dots + \frac{\partial \Delta_{itK-1}}{\partial \theta_1} \sum_{g_1=1}^K \sum_{g_2=1}^K \frac{\partial h_{ikg_1g_2}^{(t)}}{\partial \Delta_{itK-1}} P_{it-1g_1g_2}^J \\
= \frac{\partial \pi_{ik}^{(t)}}{\partial \theta_1} - \sum_{g_1=1}^K \sum_{g_2=1}^K h_{ikg_1g_2}^{(t)} \frac{\partial P_{it-1g_1g_2}^J}{\partial \theta_1}, \\
\frac{\partial \Delta_{it1}}{\partial \alpha_{ab}^{(d)}} \sum_{g_1=1}^K \sum_{g_2=1}^K \frac{\partial h_{ikg_1g_2}^{(t)}}{\partial \Delta_{it1}} P_{it-1g_1g_2}^J + \dots + \frac{\partial \Delta_{itK-1}}{\partial \alpha_{ab}^{(d)}} \sum_{g_1=1}^K \sum_{g_2=1}^K \frac{\partial h_{ikg_1g_2}^{(t)}}{\partial \Delta_{itK-1}} P_{it-1g_1g_2}^J \\
= - \sum_{g_1=1}^K \sum_{g_2=1}^K \frac{\partial h_{ikg_1g_2}^{(t)}}{\partial \gamma_{itb}^{(d)}} \frac{\partial \gamma_{itb}^{(d)}}{\partial \alpha_{ab}^{(d)}} P_{it-1g_1g_2}^J,
\end{aligned}$$

where  $h_{itkg_1g_2}^{(t)} = P(Y_{it} = k | Y_{it-1} = g_1, Y_{it-2} = g_2, x_{it})$  and  $P_{it-1g_1g_2}^J = P(Y_{it-1} = g_1, Y_{it-2} = g_2)$ .

After obtaining computations for time  $t$ , we make calculations for time  $t+1$  that



require an update  $(P_{itkl}^J, h_{ilg}^{(t)}) \rightarrow P_{it+1kl}^J$  as well as updates for derivatives,

$$\begin{aligned} P_{it+1kl}^J &= P(Y_{it+1} = k, Y_{it} = l) \\ &= \sum_{g=1}^K P(Y_{it+1} = k | Y_{it} = l, Y_{it-1} = g) P(Y_{it} = l, Y_{it-1} = g) \\ &= \sum_{g=1}^K h_{iklg}^{(t+1)} P_{itlg}^J, \end{aligned}$$

$$\begin{aligned} \frac{\partial P_{it+1kl}^J}{\partial \theta_1} &= \sum_{g=1}^K \sum_{m=1}^{K-1} \frac{\partial h_{iklg}^{(t)}}{\partial \Delta_{itm}} \frac{\partial \Delta_{itm}}{\partial \theta_1} P_{itlg}^J + \sum_{g=1}^K h_{iklg}^{(t)} \frac{\partial P_{itlg}^J}{\partial \theta_1}, \\ \frac{\partial P_{it+1kl}^J}{\partial \alpha_{ab}^{(d)}} &= \sum_{g=1}^K \sum_{m=1}^{K-1} \frac{\partial h_{iklg}^{(t)}}{\partial \Delta_{itm}} \frac{\partial \Delta_{itm}}{\partial \alpha_{ab}^{(d)}} P_{itlg}^J + \sum_{g=1}^K \frac{h_{iklg}^{(t)}}{\partial \gamma_{itab}^{(d)}} \frac{\partial \gamma_{itab}^{(d)}}{\partial \alpha_{ab}^{(d)}} P_{itlg}^J + \sum_{g=1}^K h_{iklg}^{(t)} \frac{\partial P_{itlg}^J}{\partial \alpha_{ab}^{(d)}}. \end{aligned}$$

Estimates of  $\Delta_{it}$  for OMTM(2) can be obtained using Newton-Raphson. Let

$$f(\Delta_{it}) = (f_1(\Delta_{it}), \dots, f_{K-1}(\Delta_{it})),$$

where  $f_k(\Delta_{it}) = \sum_{g_1=1}^K \sum_{g_2=1}^K h_{ikg_1g_2}^{(t)} P_{it-1g_1g_2}^J - \pi_{ik}^{(t)}$  and  $h_{itkg_1g_2}^{(t)} = P(Y_{it} = k | Y_{it-1} = g_1, Y_{it-2} = g_2, x_{it-1})$ . We obtain

$$\Delta_{it}^{(n+1)} = \Delta_{it}^{(n)} - \left( \frac{\partial f(\Delta_{it})}{\partial \Delta_{it}} \right)^{-1} f(\Delta_{it}),$$

where

$$\frac{\partial f(\Delta_{it})}{\partial \Delta_{itj}} = \begin{cases} \sum_{g_1=1}^K \sum_{g_2=1}^K h_{ikg_1g_2}^{(t)} (1 - h_{ikg_1g_2}^{(t)}) P_{it-1g_1g_2}^J, & \text{if } j = k; \\ - \sum_{g_1=1}^K \sum_{g_2=1}^K h_{ikg_1g_2}^{(t)} h_{ijg_1g_2}^{(t)} P_{it-1g_1g_2}^J, & \text{if } j \neq k. \end{cases}$$

## REFERENCES FOR WEB APPENDIX

- Barndorff-Nielsen, O.E. and Cox, D.R. (1994), *Inference and Asymptotics*, New York: Chapman and Hall.
- Cox, D.R. and Reid, N. (1987), “Parameter orthogonality and approximate conditional inference (with discussion)”, *Journal of the Royal Statistical Society, Series B*, 49, 1-39.
- Firth, D. (1987), Discussion of “Parameter orthogonality and approximate conditional inference”, *Journal of the Royal Statistical Society, Series B*, 49, 22-23.