Supplemental Materials for "Statistical Analysis of List Experiments"

Graeme Blair* Kosuke Imai†

November 3, 2011

1 The EM Algorithm for the ML Estimator of Section 2.3

In this appendix, we derive the EM algorithm for the ML estimator under the design with multiple sensitive items. Under this design, the observed-data likelihood function can be written as,

$$
\begin{aligned}
& L(\delta, \phi; Y, T, X) \\
= & \prod_{y=0}^{J} \prod_{i \in \mathcal{J}(0,y)} h(y; X_i, \phi) \prod_{t=1}^{K} \left[\prod_{i \in \mathcal{J}(t,0)} h(0; X_i, \phi)(1 - g_t(X_i, 0, \delta_{t0})) \prod_{i \in \mathcal{J}(t,J+1)} h(J; X_i, \phi) g_t(X_i, J, \delta_{tJ}) \right. \\
& \left. \times \prod_{y=1}^{J} \prod_{i \in \mathcal{J}(t,y)} \{ g_t(X_i, y-1, \delta_{t,y-1}) h(y-1; X_i, \phi) + (1 - g_t(X_i, y, \delta_{ty})) h(y; X_i, \phi) \} \right], \qquad (1)
\end{aligned}
$$

where $\mathcal{J}(t, y) = \{i : T_i = t, Y_i = y\}$ represents the set of respondents with $T_i = t$ and $Y_i = y$.

However, this likelihood function is difficult to maximize because it consists of many mixture components. Thus, following Imai (2011), we develop an EM algorithm by treating $Z^*_{i,J+t}$ as partially missing data for each $t = 1, \ldots, K$. This leads to the following complete-data likelihood function,

$$
\begin{aligned}
& L_{com}(\delta, \phi; Y, T, X, \{Z^*_{J+t}\}_{t=1}^{K}) \\
= & \prod_{i=1}^{N} \left[h(Y_i; X_i, \psi)^{\mathbf{1}\{T_i = 0\}} \prod_{t=1}^{K} \left\{ (g_t(X_i, Y_i - 1, \delta_{tY_i}) h(Y_i - 1; X_i, \psi))^{Z^*_{i,J+t}} \right. \right. \\
& \left. \left. \times (1 - g_t(X_i, Y_i, \delta_{tY_i})) h(Y_i; X_i, \psi))^{1 - Z^*_{i,J+t}} \right\}^{\mathbf{1}\{T_i = t\}} \right] \qquad (2)
\end{aligned}
$$

Given this complete-data likelihood function, the E-step of the EM algorithm is derived by computing the conditional expectation of the missing data as follows,

$$
\begin{aligned}
& \mathbb{E}(Z^*_{i,J+t} \mid Y_i = y, T_i = t, X_i = x) \\
= & \frac{\Pr(Z^*_{i,J+t} = 1, Y_i = y \mid T_i = t, X_i = x)}{\Pr(Y_i = y \mid T_i = t, X_i = x)} \qquad (3) \\
= & \begin{cases} 0 & \text{if} \quad y = 0 \\ 1 & \text{if} \quad y = J + 1 \\ \frac{g_t(x, y-1, \delta_{t,y-1}) h(y-1; x, \psi)}{g_t(x, y-1, \delta_{t,y-1}) h(y-1; x, \psi) + (1 - g_t(x, y, \delta_{ty})) h(y; x, \psi)} & \text{otherwise} \end{cases} \qquad (4)
\end{aligned}
$$

*Ph.D. candidate, Department of Politics, Princeton University, Princeton NJ 08544. Email: gblair@princeton.edu, URL: http://www.princeton.edu/~gblair

†Assistant Professor, Department of Politics, Princeton University, Princeton NJ 08544. Phone: 609–258–6601, Email: kimai@princeton.edu, URL: http://imai.princeton.edu

where the first equality follows from Bayes' rule.

Given this E-step, the M-step can be conducted by maximizing the conditional expectation of the complete-data log-likelihood function (based upon equation (27) given the observed data (Y_i, X_i, T_i) and the current values of the parameters). As in the case of the ML estimator under the standard design (Imai, 2011), this M-step reduces to the separate fitting of each model, i.e., $h(y; x, \psi)$ and $g_t(x, y, \delta_{yt})$, based on their corresponding weighted log-likelihood functions, which can be easily accomplished in standard statistical software.

2 The Asymptotic Distribution of the Two-step NLS Estimator of Section 2.4

In this appendix, we derive the asymptotic distribution of the nonlinear least squares estimator proposed in Section 2.4, assuming that $\pi_j(X_i; \theta_j)$ is the logistic function, i.e., $\pi_j(X_i; \theta_j) = \exp(X_i^\top \beta_j)/\{1 + \exp(X_i^\top \beta_j)$. We follow the standard analytical strategy outlined in Newey and McFadden (1994, Section 6) and treat the proposed two-step estimator as a method of moments estimator. In particular, the proposed estimator solves the following first order conditions with probability approaching one as the sample size tends to infinity,

$$\frac{1}{N}\sum_{i=1} \underbrace{T_i \left(Y_i - \sum_{j=1}^{J+1} \frac{\exp(X_i^\top \beta_j)}{1+\exp(X_i^\top \beta_j)} \right) \frac{\exp(X_i^\top \beta_{J+1})}{\{1+\exp(X_i^\top \beta_{J+1})\}^2} X_i}_{g(Y_i, T_i, X_i, \Theta)} = 0 \tag{5}$$

$$\frac{1}{N}\sum_{i=0} \underbrace{(1-T_i) \left(Z_{i1} - \frac{\exp(X_i^\top \beta_1)}{1+\exp(X_i^\top \beta_1)} \right) \frac{\exp(X_i^\top \beta_1)}{\{1+\exp(X_i^\top \beta_1)\}^2} X_i}_{h(Z_{i1}, T_i, X_i, \theta_1)} = 0 \tag{6}$$

$$\vdots$$

$$\frac{1}{N}\sum_{i=0} \underbrace{(1-T_i) \left(Z_{iJ} - \frac{\exp(X_i^\top \beta_J)}{1+\exp(X_i^\top \beta_J)} \right) \frac{\exp(X_i^\top \beta_J)}{\{1+\exp(X_i^\top \beta_J)\}^2} X_i}_{h(Z_{iJ}, T_i, X_i, \theta_J)} = 0 \tag{7}$$

Under the standard regularity conditions (Hansen 1982), the two-step estimator is consistent given that $\pi_j(X_i; \theta_j)$ can be consistently estimated for each control item $j = 1, \ldots, J$ using just the control group.

To derive the asymptotic variance, we utilize the sandwich robust variance formula to account for the possible correlation across control items in the control group. The asymptotic distribution is given by,

$$\sqrt{n} \begin{bmatrix} \hat{\theta}_{J+1} - \theta_{J+1} \\ \hat{\theta}_1 - \theta_1 \\ \vdots \\ \hat{\theta}_J - \theta_J \end{bmatrix} \xrightarrow{D} \mathcal{N}(0, V) \quad \text{where} \quad V = G^{-1} F G^{-1}. \tag{8}$$

The expressions for F and G are given below,

$$F = \begin{bmatrix} \mathbb{E}(g(Y_i, T_i, X_i, \Theta) g(Y_i, T_i, X_i, \Theta)^\top) & 0 & \cdots & 0 \\ 0 & \mathbb{E}(h(Z_{i1}, T_i, X_i, \theta_1) h(Z_{i1}, T_i, X_i, \theta_1)^\top) & \cdots & \mathbb{E}(h(Z_{i1}, T_i, X_i, \theta_1) h(Z_{iJ}, T_i, X_i, \theta_J)^\top) \\ \vdots & \vdots & \ddots & \vdots \\ 0 & \mathbb{E}(h(Z_{iJ}, T_i, X_i, \theta_J) h(Z_{i1}, T_i, X_i, \theta_1)^\top) & \cdots & \mathbb{E}(h(Z_{iJ}, T_i, X_i, \theta_J) h(Z_{iJ}, T_i, X_i, \theta_J)^\top) \end{bmatrix} \tag{9}$$

and

$$G^{-1} = \begin{bmatrix} H & L \\ 0 & M \end{bmatrix}^{-1} = \begin{bmatrix} H^{-1} & -H^{-1} L M^{-1} \\ 0 & M^{-1} \end{bmatrix} \tag{10}$$

where the submatrices of G are given by,

$$H = \mathbb{E}\left(\frac{\partial g(Y_i,T_i,X_i,\Theta)}{\partial \beta_{J+1}^\top}\right) = -\mathbb{E}\left(\frac{T_i \exp(2X_i^\top \beta_{J+1})}{\{1+\exp(X_i^\top \beta_{J+1})\}^4} X_i X_i^\top\right) \quad (11)$$

$$L_j = \mathbb{E}\left(\frac{\partial g(Y_i,T_i,X_i,\Theta)}{\partial \beta_j^\top}\right) = -\mathbb{E}\left(\frac{T_i \exp\{X_i^\top(\beta_j+\beta_{J+1})\}}{\{1+\exp(X_i^\top \beta_j)\}^2\{1+\exp(X_i^\top \beta_{J+1})\}^2} X_i X_i^\top\right) \quad (12)$$

$$M_{jj'} = \mathbb{E}\left(\frac{\partial h(Z_{ij},T_i,X_i,\theta_j)}{\partial \beta_{j'}^\top}\right) = \begin{cases} -\mathbb{E}\left(\frac{(1-T_i)\exp(2X_i^\top \beta_j)}{\{1+\exp(X_i^\top \beta_j)\}^4} X_i X_i^\top\right) & \text{if } j = j' \\ 0 & \text{otherwise} \end{cases} \quad (13)$$

where $L = [L_1 L_2 \ldots L_J]$ and $M_{jj'}$ is the (j,j')th block submatrix of M.

When the first step is replaced with the maximum likelihood estimation of the logistic regression model, the $h(Z_{ij},T_i,X_i,\theta_j)$ function needs to be redefined as,

$$(1-T_i)\left(Z_{ij} - \frac{\exp(X_i^\top \beta_j)}{1+\exp(X_i^\top \beta_j)}\right) X_i, \quad (14)$$

for $j = 1,\ldots,J$, while the rest of the formulae remains the same.

3 The EM Algorithm of the ML Estimator in Section 3.2

Under the setup described in Section 3.2, the likelihood function under the standard assumption given in Imai (2011) can be modified to,

$$\begin{aligned}
& L(\phi,\psi,\kappa,\delta;Y,T,X) \\
= & \prod_{y=0}^{J} \prod_{i\in\mathcal{J}(0,y)} h(y;X_i,\psi) \prod_{i\in\mathcal{J}(1,0)} \{(1-g(X_i,\delta)) + g(X_i,\delta)\underline{q}(X_i,\kappa)\} h(0;X_i,\psi) \\
\times & \prod_{i\in\mathcal{J}(1,1)} \left\{g(X_i,\delta)(1-\underline{q}(X_i,\kappa))h(0;X_i,\psi) + (1-g(X_i,\delta))h(1;X_i,\psi)\right\} \\
\times & \prod_{y=2}^{J-1} \prod_{i\in\mathcal{J}(1,y)} \{g(X_i,\delta)h(y-1;X_i,\psi) + (1-g(X_i,\delta))h(y;X_i,\psi)\} \\
\times & \prod_{i\in\mathcal{J}(1,J)} [g(X_i,\delta)\{h(J-1;X_i,\psi) + \bar{q}(X_i,\phi)h(J;X_i,\psi)\} + (1-g(X_i,\delta))h(J;X_i,\psi)] \\
\times & \prod_{i\in\mathcal{J}(1,J+1)} (1-\bar{q}(X_i,\phi)) g(X_i,\delta) h(J;X_i,\psi), \quad (15)
\end{aligned}$$

where $\mathcal{J}(t,y)$ represents a set of respondents with $(T_i,Y_i) = (t,y)$. The form of this likelihood function shows that each subset of respondents in the treatment group is a mixture of different respondent types.

To maximize this complex likelihood function, we again adopt the EM algorithm by considering $(Z_{i,J+1}^*, Z_{i,J+1}(1))$ for the respondents in the treatment group as (partially) missing data. First, the complete-data likelihood function can be written as,

$$\begin{aligned}
& L_{com}(\phi,\psi,\kappa,\delta;Z_{J+1}^*,Z_{J+1}(1),Y,T,X) \\
= & \prod_{i=1}^{N} h(Y_i;X_i,\psi)^{1-T_i} \left\{h(Y_i;X_i,\psi)(1-g(X_i,\delta))\right\}^{T_i(1-Z_{i,J+1}^*)} \\
& \times \Big[g(X_i,\delta) h(Y_i-1;X_i,\psi)^{Z_{i,J+1}(1)} h(Y_i;X_i,\psi)^{1-Z_{i,J+1}(1)} \bar{q}(X_i,\phi)^{(1-Z_{i,J+1}(1))\mathbf{1}\{Y_i=J\}} \\
& \quad (1-\bar{q}(X_i,\phi))^{Z_{i,J+1}(1)\mathbf{1}\{Y_i=J+1\}} \underline{q}(X_i,\kappa)^{(1-Z_{i,J+1}(1))\mathbf{1}\{Y_i=0\}} (1-\underline{q}(X_i,\kappa))^{Z_{i,J+1}(1)\mathbf{1}\{Y_i=1\}} \Big]^{T_i Z_{i,J+1}^*} \quad (16)
\end{aligned}$$

Then, the E-step of the EM algorithm requires the calculation of the following conditional expectations,

$$\mathbb{E}(Z^*_{i,J+1} \mid Y_i = y, T_i = 1, X_i = x) = \begin{cases} \frac{h(0;x,\psi)g(x,\delta)\underline{q}(x,\kappa)}{h(0;x,\psi)\{g(x,\delta)\underline{q}(x,\kappa)+(1-g(x,\delta))\}} & \text{if } y = 0 \\ \frac{h(0;x,\psi)g(x,\delta)(1-\underline{q}(x,\kappa))}{h(0;x,\psi)g(x,\delta)(1-\underline{q}(x,\kappa))+h(1;x,\psi)(1-g(x,\delta))} & \text{if } y = 1 \\ \frac{g(x,\delta)\{h(J-1;x,\psi)+\bar{q}(x,\phi)h(J;x,\psi)\}}{g(x,\delta)\{h(J-1;x,\psi)+\bar{q}(x,\phi)h(J;x,\psi)\}+(1-g(x,\delta))h(J;x,\psi)} & \text{if } y = J \\ 1 & \text{if } y = J+1 \\ \frac{g(x,\delta)h(y-1;x,\psi)}{g(x,\delta)h(y-1;x,\psi)+(1-g(x,\delta))h(y;x,\psi)} & \text{otherwise} \end{cases} \tag{17}$$

$$\mathbb{E}(Z^*_{i,J+1}Z_{i,J+1}(1) \mid Y_i = y, T_i = 1, X = x) = \begin{cases} 0 & \text{if } y = 0 \\ \frac{g(x,\delta)h(J-1;x,\psi)}{g(x,\delta)\{h(J-1;x,\psi)+\bar{q}(x,\phi)h(J;x,\psi)\}+(1-g(x,\delta))h(J;x,\psi)} & \text{if } y = J \\ \mathbb{E}(Z^*_{i,J+1} \mid Y = y, T = 1, X = x) & \text{otherwise} \end{cases} \tag{18}$$

The M-step of the EM algorithm consists of the separate fitting of each component of the model, i.e., $g(x, \delta), h(y; x, \psi), \bar{q}(x, \phi)$, and $\underline{q}(x, \kappa)$, based on the weighted log-likelihood function, for which the appropriate weights are calculated at each iteration based on the above conditional expectations.

References

Hansen, L. P. (1982). Large sample properties of generalized method of moments estimators. *Econometrica*, **50**(4), 1029–1054.

Imai, K. (2011). Multivariate regression analysis for the item count technique. *Journal of the American Statistical Association*, **106**(494), 407–416.

Newey, W. and McFadden, D. (1994). *Handbook of Econometrics (eds. R. F. Engle and D. L. McFadden)*, volume IV, chapter Large Sample Estimation and Hypothesis Testing, pages 2111–2245. North Holland, Amsterdam.