Technical Report for

"Consumer Shopping and Spending Across Retail Formats"

by

Edward J. Fox, Alan L. Montgomery and Leonard M. Lodish

Markov Chain Monte Carlo Estimation of a Hierarchical Multivariate Type-2 Tobit Model

We begin by modifying our notation. We rewrite the parameters for equation (2), the expenditure equation, $\boldsymbol{\omega}_{hi} = [\ \alpha_{hi}\ \ \boldsymbol{\beta}'_i\ \ \gamma_i\ \ \boldsymbol{\delta}'_i\ ]'$. Similarly, we rewrite the parameters for equation (4), the patronage equation, $\boldsymbol{\zeta}_{hi} = [\ \kappa_{hi}\ \ \boldsymbol{\lambda}'_i\ \ \rho_i\ \ \boldsymbol{\psi}'_i\ ]'$. We also rewrite the predictor variables common to the two equations, $\mathbf{m}_{hit} = [\ 1\ \ \mathbf{x}'_{hit}\ \ t_{hi}\ \ \mathbf{s}'_t\ ]'$. We then stack (1) the dependent variables of both equations for all households h and time periods t so that $\mathbf{y}^*_i = [\ y^*_{1i1}\ \ y^*_{1i2}\ \dots\ y^*_{HiT}\ ]'$ and $\mathbf{z}^*_i = [\ z^*_{1i1}\ \ z^*_{1i2}\ \dots\ z^*_{HiT}\ ]'$, (2) the error terms of both equations for all households h and time periods t so that $\boldsymbol{\varepsilon}_i = [\ \varepsilon_{1i1}\ \ \varepsilon_{1i2}\ \dots\ \varepsilon_{HiT}\ ]'$ and $\mathbf{u}_i = [\ u_{1i1}\ \ u_{1i2}\ \dots\ u_{HiT}\ ]'$, and finally (3) the predictor variables shared by the two equations, $M_i = [\ \mathbf{m}_{1i1}\ \ \mathbf{m}_{1i2}\ \dots\ \mathbf{m}_{HiT}\ ]'$. We allow for contemporaneous correlation of the error terms in equations (2) and (4) by adopting the SUR forms shown below.

$$\begin{bmatrix}\mathbf{y}_1^*\\ \mathbf{y}_2^*\\ \vdots\\ \mathbf{y}_S^*\end{bmatrix} = \begin{bmatrix}M_1 & & & \\ & M_2 & & \\ & & \ddots & \\ & & & M_S\end{bmatrix}\begin{bmatrix}\boldsymbol{\omega}_1\\ \boldsymbol{\omega}_2\\ \vdots\\ \boldsymbol{\omega}_S\end{bmatrix} + \begin{bmatrix}\boldsymbol{\varepsilon}_1\\ \boldsymbol{\varepsilon}_2\\ \vdots\\ \boldsymbol{\varepsilon}_S\end{bmatrix} \quad , \quad \begin{bmatrix}\mathbf{z}_1^*\\ \mathbf{z}_2^*\\ \vdots\\ \mathbf{z}_S^*\end{bmatrix} = \begin{bmatrix}M_1 & & & \\ & M_2 & & \\ & & \ddots & \\ & & & M_S\end{bmatrix}\begin{bmatrix}\boldsymbol{\zeta}_1\\ \boldsymbol{\zeta}_2\\ \vdots\\ \boldsymbol{\zeta}_S\end{bmatrix} + \begin{bmatrix}\mathbf{u}_1\\ \mathbf{u}_2\\ \vdots\\ \mathbf{u}_S\end{bmatrix}$$

where, for $\{i = 1, ..., S\}$:

$\boldsymbol{\varepsilon}_i$ is an HT vector of disturbances such that $E(\boldsymbol{\varepsilon}_i) = 0$ and $E(\boldsymbol{\varepsilon}_i\boldsymbol{\varepsilon}_j') = \sigma_{ij}I_{HT}$, and

$\mathbf{u}_i$ is an HT vector of disturbances such that $E(\mathbf{u}_i) = 0$ and $E(\mathbf{u}_i\mathbf{u}_j') = \lambda_{ij}I_{HT}$,

$$\text{with } \Sigma = \begin{bmatrix}\sigma_{11} & \sigma_{12} & \cdots & \sigma_{1S}\\ \sigma_{21} & \sigma_{22} & \cdots & \sigma_{2S}\\ \vdots & \vdots & \ddots & \vdots\\ \sigma_{S1} & \sigma_{S2} & \cdots & \sigma_{SS}\end{bmatrix} \quad , \quad \Lambda = \begin{bmatrix}\theta_{11} & \lambda_{12} & \cdots & \lambda_{1S}\\ \lambda_{21} & \lambda_{22} & \cdots & \lambda_{2S}\\ \vdots & \vdots & \ddots & \vdots\\ \lambda_{S1} & \lambda_{S2} & \cdots & \lambda_{SS}\end{bmatrix}$$

Finally, for clarity we rewrite the SUR equations above as: $\mathbf{y}^* = M\boldsymbol{\omega} + \boldsymbol{\varepsilon}$ and $\mathbf{z}^* = M\boldsymbol{\zeta} + \mathbf{u}$.

Now, we write the hierarchies associated with the two shopping decisions. We stack (1) the intercept coefficients for all households h so that $\boldsymbol{\alpha}_i = [\ \alpha_{1i}\ \ \alpha_{2i}\ \dots\ \alpha_{Hi}\ ]'$ and $\boldsymbol{\kappa}_i = [\ \kappa_{1i}\ \ \kappa_{2i}\ \dots\ \kappa_{Hi}\ ]'$, (2) the common predictor variables for the two equations, $W = [\ \mathbf{w}_1\ \ \mathbf{w}_2\ \dots\ \mathbf{w}_H\ ]'$, and (3) the error terms of the hierarchical equations for all households h, $\boldsymbol{\eta}_i = [\ \eta_{1i}\ \ \eta_{2i}\ \dots\ \eta_{Hi}\ ]'$ and $\boldsymbol{\nu}_i = [\ \nu_{1i}\ \ \nu_{2i}\ \dots\ \nu_{Hi}\ ]'$. We allow for contemporaneous correlation of unexplained preferences across store chains in equations (5) and (6) by using the SUR forms below.

$$\begin{bmatrix}\boldsymbol{\alpha}_1\\ \boldsymbol{\alpha}_2\\ \vdots\\ \boldsymbol{\alpha}_S\end{bmatrix}=\begin{bmatrix}W & & & \\ & W & & \\ & & \ddots & \\ & & & W\end{bmatrix}\begin{bmatrix}\boldsymbol{\delta}_1\\ \boldsymbol{\delta}_2\\ \vdots\\ \boldsymbol{\delta}_S\end{bmatrix}+\begin{bmatrix}\boldsymbol{\xi}_1\\ \boldsymbol{\xi}_2\\ \vdots\\ \boldsymbol{\xi}_S\end{bmatrix}\quad,\quad\begin{bmatrix}\boldsymbol{\iota}_1\\ \boldsymbol{\iota}_2\\ \vdots\\ \boldsymbol{\iota}_S\end{bmatrix}=\begin{bmatrix}W & & & \\ & W & & \\ & & \ddots & \\ & & & W\end{bmatrix}\begin{bmatrix}\boldsymbol{\psi}_1\\ \boldsymbol{\psi}_2\\ \vdots\\ \boldsymbol{\psi}_S\end{bmatrix}+\begin{bmatrix}\boldsymbol{\tau}_1\\ \boldsymbol{\tau}_2\\ \vdots\\ \boldsymbol{\tau}_S\end{bmatrix}$$

where, for $\{i = 1, ..., S\}$ $\boldsymbol{\xi}_i$ is an H vector of disturbances such that $E(\boldsymbol{\xi}_i) = 0$ and $E(\boldsymbol{\xi}_i\boldsymbol{\xi}_j') = v^{\alpha}_{ij}I_H$, and $\boldsymbol{\tau}_i$ is an H vector of disturbances such that $E(\boldsymbol{\tau}_i) = 0$ and $E(\boldsymbol{\tau}_i\boldsymbol{\tau}_j') = v^{\iota}_{ij}I_H$,

$$\text{with } V_\alpha=\begin{bmatrix}v^\alpha_{11} & v^\alpha_{12} & \cdots & v^\alpha_{1S}\\ v^\alpha_{21} & v^\alpha_{22} & \cdots & v^\alpha_{2S}\\ \vdots & \vdots & \ddots & \vdots\\ v^\alpha_{S1} & v^\alpha_{S2} & \cdots & v^\alpha_{SS}\end{bmatrix}\quad,\quad V_\iota=\begin{bmatrix}v^\iota_{11} & v^\iota_{12} & \cdots & v^\iota_{1S}\\ v^\iota_{21} & v^\iota_{22} & \cdots & v^\iota_{2S}\\ \vdots & \vdots & \ddots & \vdots\\ v^\iota_{S1} & v^\iota_{S2} & \cdots & v^\iota_{SS}\end{bmatrix}$$

We summarize the SUR equations above as: $\boldsymbol{\alpha} = \mathrm{W}\boldsymbol{\delta} + \boldsymbol{\xi}$ and $\boldsymbol{\iota} = W\boldsymbol{\psi} + \boldsymbol{\tau}$.

Both the SUR structures of the models above and the hierarchical specification preclude analytical solutions of the S-model system of equations under consideration. Moreover, the high dimension of the integral makes the use of numerical integration techniques infeasible for our systems of equations. Due to the limitations of analytical and numerical estimation techniques for the hierarchical multivariate Tobit specification, we use an MCMC approach to estimate the marginal distributions of the latent dependent variables, parameters and covariances. The MCMC algorithm involves sampling sequentially from the relevant conditional distributions over a large number of iterations. These draws can be shown to converge to the marginal posterior distributions.

Our implementation of the MCMC algorithm has three steps that are described below.

A. Conditional distributions

The first implementation step requires that we specify conditional distributions of the relevant variables. The solutions of these distributions follow from the normality assumption of the disturbance terms. We employ natural conjugate priors. Specifications of the conditional distributions are as follows:

1. y^*_{hit} is y_{hit} if y_{hit}>0, otherwise y^*_{hit} is drawn from a normal distribution, truncated above at 0.

$$y^*_{hit}|\mathbf{y}^*_{h,j\neq i,t},\boldsymbol{\omega}_i,\alpha_{hi},\Sigma\sim\begin{cases} y_{hit}|y_{hit}>0\\ N_T\left(\mathbf{m}_{hit}\boldsymbol{\omega}_i+\alpha_{hi}-\boldsymbol{\sigma}_{ij}\Sigma^{-1}_{jj}\mathbf{y}_{h,j\neq i,t},\sigma_{ii}-\boldsymbol{\sigma}_{ij}\Sigma^{-1}_{jj}\boldsymbol{\sigma}_{ji}\right) \textit{otherwise}\end{cases}$$

$$\text{where: }\mathbf{y}^*_{ht}=\begin{bmatrix}y^*_{hit}\\ --\\ \mathbf{y}^*_{h,j\neq i,t}\end{bmatrix}\text{ and }\boldsymbol{\Sigma}=\begin{bmatrix}\sigma_{ii} & | & \boldsymbol{\sigma}_{ij}\\ -- & + & --\\ \boldsymbol{\sigma}_{ji} & | & \Sigma_{jj}\end{bmatrix}$$

As the notation suggests, the $\mathbf{y}^*_{ht}$ vector and **G** matrix are partitioned between the store chain of interest, *i*, and all other store chains, *j* ... *i*. Without loss of generality, we have shown the store chain of interest to be the first. Each chain is then drawn in succession for household *h*, conditioning on $\mathbf{y}^*_{h,i\neq j,t}$, a vector of latent dependent variables for all *j* ... *i*, and **G**.

The truncated normal variates are drawn using the inverse cdf method. Given the truncation value of zero, the conditional expected value of the dependent variable,

$$E\left(y^{*(t)}_{hit}\right)|\mathbf{y}^{*(t-1)}_{h,j\neq i,t},\boldsymbol{\omega}^{(t-1)}_{i},\alpha^{(t-1)}_{hi},\Sigma^{(t-1)} = \mathbf{m}_{hit}\boldsymbol{\omega}^{(t-1)}_{i} + \alpha^{(t-1)}_{hi} - \sigma^{(t-1)}_{ij}\Sigma^{-1(t-1)}_{jj}\left(\mathbf{y}^{*(t-1)}_{h,j\neq i,t} - E\left(\mathbf{y}^{*(t-1)}_{h,j\neq i,t}\right)\right),$$

and the conditional standard deviation of the dependent variable,

$\sigma^{(t-1)}_{y^*}|\Sigma^{(t-1)} = \sigma^{(t-1)}_{ii} - \boldsymbol{\sigma}^{(t-1)}_{ij}\Sigma^{-1(t-1)}_{jj}\boldsymbol{\sigma}^{(t-1)}_{ji}$, we draw the truncated normal values using the following procedure. (Note that conditioning arguments are dropped for clarity):

- Compute the upper limit for uniform interval: $L = \Phi\left[\left(0 - E\left(y^{*(t)}_{hit}\right)\right)/\sigma^{(t-1)}_{y^*}\right]$, where **M**[] represents the Normal cdf.
- Draw a uniform variate: U ~ *Uniform*(0,L).
- Compute the realized value of the uniform draw: $y^{*(t)}_{hit} = \Phi^{-1}(U)\sigma^{(t-1)}_{y^*} + E\left(y^{*(t)}_{hit}\right)$.

Note that, when using this procedure, values of U approaching 0 tend toward - 4 while values of U approaching L tend toward zero, the truncation point.

2. In a similar fashion, we draw the latent dependent variable values for the probit component of the model. If the indicator variable $z_{hit} = 1$, then z^*_{hit} is drawn from a normal distribution, truncated below at 0. Otherwise, z^*_{hit} is drawn from a normal distribution, truncated above at 0.

$$z^*_{hit}|\mathbf{z}^*_{h,j\neq i,t},\boldsymbol{\zeta}_i,\iota_{hi},\Lambda \sim N_T\left(\mathbf{m}_{hit}\boldsymbol{\zeta}_i + \iota_{hi} - \boldsymbol{\lambda}_{ij}\Lambda^{-1}_{jj}\mathbf{z}^*_{h,j\neq i,t},\lambda_{ii} - \boldsymbol{\lambda}_{ij}\Lambda^{-1}_{jj}\boldsymbol{\lambda}_{ji}\right)$$

where: $\mathbf{z}^*_{ht} = \begin{bmatrix} z^*_{hit} \\ -- \\ \mathbf{z}^*_{h,j\neq i,t} \end{bmatrix}$ and $\Lambda = \begin{bmatrix} \lambda_{ii} & | & \boldsymbol{\lambda}_{ij} \\ -- & + & -- \\ \boldsymbol{\lambda}_{ji} & | & \Lambda_{jj} \end{bmatrix}$

As in the conditional spending model, the latent probit dependent variables are drawn using the inverse cdf method with mean and variance as follows:

$$E\left(z_{hit}^{*(t)}\right)|\mathbf{z}_{h,j\neq i,t}^{*(t-1)},\zeta_i^{(t-1)},\iota_{hi}^{(t-1)},\Lambda^{(t-1)} = \mathbf{m}_{hit}\zeta_i^{(t-1)} + \iota_{hi}^{(t-1)} - \lambda_{ij}^{(t-1)}\Lambda_{jj}^{-1(t-1)}\left(\mathbf{z}_{h,j\neq i,t}^{*(t-1)} - E\left(\mathbf{z}_{h,j\neq i,t}^{*(t-1)}\right)\right)$$

$$\lambda_{y^*}^{(t-1)}|\Lambda^{(t-1)} = \lambda_{ii}^{(t-1)} - \boldsymbol{\lambda}_{ij}^{(t-1)}\Lambda_{jj}^{-1(t-1)}\boldsymbol{\lambda}_{ji}^{(t-1)}$$

3. The store-level parameters in **T**$_{hi}$ (which include a mean intercept term, $\overline{\alpha_i}$) are drawn from a SUR model with variance/covariance matrix of disturbances **G**. Individual intercepts are drawn in step 5.

$$\boldsymbol{\omega}^{(t)}|\mathbf{y}^{*(t)},\Sigma^{(t-1)} \sim N\left(O\left(M_i'\left(\Sigma^{-1(t-1)}\otimes I_{HT}\right)\mathbf{y}_{\mathbf{i}}^{*(t)}\right),O\right)$$

where $O = \left(M_i'\left(\Sigma^{-1(t-1)}\otimes I_{HT}\right)M_i\right)^{-1}$

4. The store-level parameters in ▪$_{hi}$ (which include a mean intercept term, $\bar{\iota}_i$) are drawn from a SUR model with variance/covariance matrix of disturbances **7**. Individual intercepts are drawn in step 6.

$$\boldsymbol{\varsigma}^{(t)}|\mathbf{z}^{*(t)},\Lambda^{(t-1)} \sim N\left(P\left(M_i'\left(\Lambda^{-1(t-1)}\otimes I_{HT}\right)\mathbf{z}_i^{*(t)}\right),P\right)$$

where $P = \left(M_i'\left(\Lambda^{-1(t-1)}\otimes I_{HT}\right)M_i\right)^{-1}$

5. The vector of household intercepts "$_h$ is drawn from a SUR model with variance/covariance matrix of disturbances **G**.

$$\boldsymbol{\alpha}_h^{(t)}|\mathbf{y}^{*(t)},\boldsymbol{\omega}^{(t)},\Sigma^{(t-1)},V_\alpha^{(t-1)}\boldsymbol{\delta}^{(t-1)},\overline{\boldsymbol{\alpha}}_h \sim N\left(Q\left(U_h'\left(\Sigma^{-1(t-1)}\otimes I_T\right)\mathbf{r}_h^\alpha + V_a^{-1(t-1)}W_h\boldsymbol{\delta}^{(t-1)}\right),Q\right)$$

where $Q = \left(U_h'\left(\Sigma^{-1(t-1)}\otimes I_T\right)U_h + V_\alpha^{(t-1)-1}\right)^{-1}$,

$$U = \begin{bmatrix} \mathbf{1}_T & & & \\ & \mathbf{1}_T & & \\ & & \ddots & \\ & & & \mathbf{1}_T \end{bmatrix},\quad W_h = \begin{bmatrix} \mathbf{w}_h' & & & \\ & \mathbf{w}_h' & & \\ & & \ddots & \\ & & & \mathbf{w}_h' \end{bmatrix}, \text{ and}$$

$$\mathbf{r}_h^\alpha = \begin{bmatrix} \mathbf{y}_{h1}^{*(t)} \\ \mathbf{y}_{h2}^{*(t)} \\ \vdots \\ \mathbf{y}_{hS}^{*(t)} \end{bmatrix} - \begin{bmatrix} M_{h1} & & & \\ & M_{h2} & & \\ & & \ddots & \\ & & & M_{hS} \end{bmatrix}\boldsymbol{\omega}^{(t)}.$$

6. The vector of household intercepts **4**$_h$ is also drawn from a SUR model with variance/covariance matrix of disturbances **7**.

$$\boldsymbol{\iota}_h^{(t)} | \mathbf{z}^{*(t)}, \boldsymbol{\zeta}^{(t)}, \Lambda^{(t-1)}, V_\iota^{(t-1)}, \boldsymbol{\psi}^{(t-1)}, \bar{\boldsymbol{\iota}}_h \sim N\left(R\left(U_h'\left(\Lambda^{-1(t-1)} \otimes I_T\right)\mathbf{r}_h^\iota + V_\iota^{-1(t-1)} W_h \boldsymbol{\psi}^{(t-1)}\right), R\right)$$

where $R = \left(U_h'\left(\Lambda^{-1(t-1)} \otimes I_T\right)U_h + V_\iota^{(t-1)-1}\right)^{-1}$, and

$$\mathbf{r}_h^\iota = \begin{bmatrix} \mathbf{z}_{h1}^{*(t)} \\ \mathbf{z}_{h2}^{*(t)} \\ \vdots \\ \mathbf{z}_{hS}^{*(t)} \end{bmatrix} - \begin{bmatrix} M_{h1} & & & \\ & M_{h2} & & \\ & & \ddots & \\ & & & M_{hS} \end{bmatrix} \varsigma^{(t)}.$$

7. The vector of hyper-parameters, *****, is drawn from a SUR model with variance/covariance matrix of disturbances, *V*".

$$\boldsymbol{\delta}^{(t)} | \boldsymbol{\alpha}^{(t)}, V_\alpha^{(t-1)}, V_\delta, \bar{\boldsymbol{\delta}} \sim N\left(S\left(W'\left(V_\alpha^{-1(t-1)} \otimes I_H\right)\boldsymbol{\alpha}^{(t)} + V_\delta^{-1}\bar{\boldsymbol{\delta}}\right), S\right)$$

$$\text{where } S = \left(Q'\left(V_\alpha^{-1(t-1)} \otimes I_H\right)Q + V_\delta^{-1}\right)^{-1}, \text{ and } Q = \begin{bmatrix} D & & & \\ & D & & \\ & & \ddots & \\ & & & D \end{bmatrix}$$

8. The vector of hyper-parameters, **R**, is drawn from a SUR model with variance/covariance matrix of disturbances, *V* **4**.

$$\boldsymbol{\psi}^{(t)} | \boldsymbol{\iota}^{(t)}, V_\iota^{(t-1)}, V_\psi, \bar{\boldsymbol{\psi}} \sim N\left(T\left(W'\left(V_\iota^{-1(t-1)} \otimes I_H\right)\boldsymbol{\iota}^{(t)} + V_\psi^{-1}\bar{\boldsymbol{\psi}}\right), T\right)$$

$$\text{where } T = \left(Q'\left(V_\iota^{-1(t-1)} \otimes I_H\right)Q + V_\psi^{-1}\right)^{-1}$$

9. **G** is drawn from an inverted Wishart distribution with *HT*+**<**$_{\mathbf{E}}$ degrees of freedom.

$$\Sigma^{-1(t)} | \boldsymbol{\omega}^{(t)}, \mathbf{y}^{*(t)}, \boldsymbol{\delta}^{(t)}, \Sigma^{(t)}, V_\alpha^{(t)}, V_\Sigma, v_\Sigma \sim W\left(HT + v_\Sigma, \left(V_\Sigma + \boldsymbol{\varepsilon\varepsilon}'\right)^{-1}\right)$$

10. **7** is also drawn from an inverted Wishart distribution with *HT*+**<**$_{\mathbf{7}}$ degrees of freedom.

$$\Lambda^{-1(t)}|\boldsymbol{\zeta}^{(t)},\mathbf{z}^{*(t)},\boldsymbol{\psi}^{(t)},\Lambda^{(t)},V_\iota^{(t)},V_\Lambda,\nu_\Lambda \sim W\left(HT+\nu_\Lambda,\left(V_\Lambda+\mathbf{uu}'\right)^{-1}\right)$$

11. *V*" is drawn from an inverted Wishart distribution with *H*+<" degrees of freedom.

$$V_\alpha^{-1(t)}|\boldsymbol{\omega}^{(t)},\mathbf{y}^{*(t)},\boldsymbol{\delta}^{(t)},V_\alpha^{(t)},\overline{V}_\alpha,\nu_\alpha \sim W\left(H+\nu_\alpha,\left(\overline{V}_\alpha+\boldsymbol{\xi\xi}'\right)^{-1}\right)$$

12. *V*4 is drawn from an inverted Wishart distribution with *H*+<4 degrees of freedom.

$$V_\iota^{-1(t)}|\boldsymbol{\varsigma}^{(t)},\mathbf{z}^{*(t)},\boldsymbol{\psi}^{(t)},V_\iota^{(t)},\overline{V}_\iota,\nu_\iota \sim W\left(H+\nu_\iota,\left(\overline{V}_\iota+\boldsymbol{\tau\tau}'\right)^{-1}\right)$$

B. Prior distributions

The second implementation step is to specify prior distributions for the parameters of interest. Note that the priors are set to be non-informative so that inferences are driven by the data.

1. The prior distribution of * is MVN(*,*V**), where * = **0** and *V** = diag(10^3).
2. The prior distribution of **R** is MVN(**R**,*V*R), where **R** = **0** and *V*R = diag(10^3).
3. The prior distribution of **G**$^{-1}$ is Wishart: W(<G,*V*E), where <G = 10 and *V*E = diag(10^{-3}).
4. The prior distribution of **7**$^{-1}$ is Wishart: W(<7,*V*7), where <7 = 10 and *V*7 = diag(10^{-3}).
5. The prior distribution of *V*"$^{-1}$ is Wishart: W(<",*V*"), where <" = 1 and *V*" = diag(10^{-3}).
6. The prior distribution of *V*4^{-1} is Wishart: W(<4,*V*4), where <4 = 1 and *V*4 = diag(10^{-3}).

C. Initial values

The third implementation step is to set initial values for the parameters of the marginal distributions. The starting values for **T**$_i$ from equation (2) are computed by OLS, using ln(y_{hit}) as the dependent variable of the regression. The individual-level intercepts, "$_{hi}$, are computed using the residuals from the regression model above as the dependent variable, and regressing the design vector *U* on those residuals using OLS. The covariance matrix, **G**, is initiated by taking the residuals of the second-stage regression,

g_{hoi}, (conditioned on the initial parameter values) and using them to compute sample covariances. In a similar fashion, the starting values for the patronage equation parameters, H_i, are computed by OLS, using z_{hit} as the dependent variable. Again, the individual-level intercepts, 4_{hi}, are computed using the residuals from the first-stage regression model as the dependent variable, and regressing the design vector U on those residuals. Again, the residuals from this second-stage regression, u_{hit}, are used to compute the sample covariances, which serve as the initial value for 7. Note that other initial values were used to ensure that estimates were not dependent on a particular starting point.

The final step is to generate $N_1 + N_2$ random draws from the conditional distributions. The number of initialization iterations, N_1, is determined empirically. We use a "burn in" period of 3500 iterations. To reduce autocorrelation in the MCMC draws, we "thin the line," using every fifth draw in the sequence that comprises N_2 for our estimation. In this way, the last N_2 iterations are used to estimate marginal posterior distributions of the parameters of interest. Note that the means and variances of these distributions are computed directly using the means and variances of the final N_2 draws of each parameter.