Supplementary Appendix to "A Social Interactions Model with Endogenous Friendship Formation and Selectivity"

Chih-Sheng Hsieh*
Department of Economics
Chinese University of Hong Kong

Lung Fei Lee†
Department of Economics
The Ohio State University

This document provides supplementary appendices for "A Social Interactions Model with Endogenous Friendship Formation and Selectivity" which is prepared for *Journal of Applied Econometrics*.

*Department of Economics, The Chinese University of Hong Kong, 9/F, Esther Lee Building, CUHK, Shatin. Tel: 3943-8006. Fax: 2603-5805. Email: cshsieh@cuhk.edu.hk.

†Department of Economics, Ohio State University, 1945 N. High St., Columbus OH 43210.

Appendix I. Posterior analysis of the network and SCSAR models

Here, we list the set of conditional posterior distributions required by the Gibbs sampler:

(i) By Bayes' theorem, $P(z_{i,g}|Y_g, W_g, Z_{-i,g}, \theta, \alpha_g) \propto \pi(z_{i,g}) \cdot P(Y_g, W_g|Z_g, \theta, \alpha_g)$, $i = 1, \cdots, m_g$, $g = 1, \cdots, G$. Therefore, we have

$$P(z_{i,g}|Y_g, W_g, Z_{-i,g}, \theta, \alpha_g) \propto \mathcal{N}_{\bar{d}}(z_{i,g}; 0, I_{\bar{d}}) \cdot P(Y_g, W_g|Z_g, \theta, \alpha_g), \tag{1}$$

where

$$\begin{aligned} &P(Y_g, W_g|Z_g, \theta, \alpha_g) \\ &= P(Y_g|W_g, Z_g, \theta, \alpha_g) \cdot P(W_g|Z_g, \theta, \alpha_g) \\ &= (2\pi)^{-\frac{m_g}{2}} \left(\sigma_u^2\right)^{-\frac{m_g}{2}} |I_{m_g} - \lambda W_g| \exp\left(-\frac{1}{2\sigma_u^2} u_g' u_g\right) \cdot \prod_{i \neq j} \frac{\exp(w_{ij,g}\psi_{ij,g})}{1 + \exp(\psi_{ij,g})}. \end{aligned}$$

(ii) We can simplify the conditional posterior distribution of ϕ to $P(\phi|\{W_g\}, \{Z_g\})$. Using Bayes' theorem, we have

$$\begin{aligned} &P(\phi|\{W_g\}, \{Z_g\}) \\ &\propto \pi(\phi) \cdot \prod_{g=1}^{G} P(W_g|Z_g, \phi) = \mathcal{N}_{2\bar{s}+\bar{q}+\bar{d}}(\phi; \phi_0, \Phi_0) \cdot \prod_{g=1}^{G} P(W_g|Z_g, \phi), \quad \phi \in O_1 \end{aligned} \tag{2}$$

(iii) By applying Bayes' theorem, we have

$$P(\lambda|\{Y_g\}, \{W_g\}, \{Z_g\}, \beta, \sigma, \{\alpha_g\}) \propto \prod_{g=1}^{G} P(Y_g|W_g, Z_g, \lambda, \beta, \sigma, \alpha_g), \tag{3}$$

where $\lambda \in [-1/\tau_G, 1/\tau_G]$.

(iv) By applying Bayes' theorem, we have

$$\begin{aligned} &P(\beta|\{Y_g\}, \{W_g\}, \{Z_g\}, \lambda, \sigma, \{\alpha_g\}) \\ &\qquad \propto \mathcal{N}_{2k}(\beta; \beta_0, B_0) \cdot \prod_{g=1}^{G} P(Y_g|W_g, Z_g, \lambda, \beta, \sigma, \alpha_g). \end{aligned}$$

Since both $\mathscr{N}_{2k}(\beta; \beta_0, B_0)$ and $P(Y_g|W_g, Z_g, \lambda, \beta, \sigma, \alpha_g)$ are in terms of normal density, we obtain standard linear model results in which

$$P(\beta|\{Y_g\}, \{W_g\}, \{Z_g\}, \lambda, \sigma, \{\alpha_g\}) \propto \mathscr{N}_{2k}\left(\beta; \hat{\beta}, \mathbf{B}\right), \tag{4}$$

where $\hat{\beta} = \mathbf{B}\left(B_0^{-1}\beta_0 + \frac{1}{\sigma_u^2}\sum_{g=1}^{G}\mathbf{X}_g'((I_{m_g} - \lambda W_g)Y_g - Z_g\sigma_{z\epsilon} - l_g\alpha_g)\right)$ and $\mathbf{B} = \left(B_0^{-1} + \frac{1}{\sigma_u^2}\sum_{g=1}^{G}\mathbf{X}_g'\mathbf{X}_g\right)^{-1}$ with $\mathbf{X}_g = (X_g, W_gX_g)$.

(v) By applying Bayes' theorem, we have

$$\begin{aligned} &P(\sigma|\{Y_g\}, \{W_g\}, \{Z_g\}, \lambda, \beta, \{\alpha_g\}) \\ &\qquad \propto \mathscr{N}_{\bar{d}+1}(\sigma; \sigma_0, \Sigma_0)\prod_{g=1}^{G}P(Y_g|W_g, X_g, Z_g, \lambda, \beta, \sigma, \alpha_g), \quad \sigma \in O_2. \end{aligned} \tag{5}$$

(vi) By applying Bayes' theorem, we have

$$P(\alpha_g|Y_g, W_g, Z_g, \lambda, \beta, \sigma) \propto \mathscr{N}(\alpha_g, \alpha_0, A_0) \cdot P(Y_g|W_g, X_g, Z_g, \lambda, \beta, \sigma, \alpha_g),$$

for $g = 1, \cdots, G$. Similar to (iv), we can further obtain

$$P(\alpha_g|Y_g, W_g, Z_g, \lambda, \beta, \sigma) \propto \mathscr{N}(\alpha_g; \hat{\alpha}_g, R_g), \tag{6}$$

where $\hat{\alpha}_g = R_g\left(A_0^{-1}\alpha_0 + \frac{1}{\sigma_u^2}l_g'\left((I_{m_g} - \lambda W_g)Y_g - \mathbf{X}_g\beta - Z_g\sigma_{z\epsilon}\right)\right)$ and $R_g = \left(A_0^{-1} + \frac{1}{\sigma_u^2}l_gl_g'\right)^{-1}$.

The sampling of β and $\{\alpha_g\}$ are straightforward because of their well-known conditional posterior distributions. However, other conditional posterior distributions are not available in closed forms, and hence, we need to use the Metropolis-Hastings (M-H) algorithm to draw from those conditional distributions. Tierney (1994) and Chib and Greenberg (1996) show that the combination of Markov chains (Metropolis-within-Gibbs) is still a Markov chain with the invariant distribution equal to the correct objective distribution. The procedure of the MCMC sampling starts with arbitrary initial values for $\{Z_g^{(0)}\}$, $\{\alpha_g^{(0)}\}$, and parameters $\theta^{(0)}$, and then the sampling proceeds sequentially from the above set of conditional posterior distributions. The implementation details of the MCMC procedure are listed as follows:

At the t^{th} iteration of the MCMC sampling, we implement the following steps:

Step 1. Sample $z_{i,g}^{(t)}$ from $P(z_{i,g}|Y_g, W_g, \theta^{(t-1)}, \alpha_g^{(t-1)})$, as specified in Eq. (1), by using the M-H algorithm for $i = 1, \cdots, m_g$ and $g = 1, \cdots, G$.

(a) Propose $\tilde{z}_{i,g} \sim \mathcal{N}_{\bar{d}}(z_{i,g}^{(t-1)}, \kappa_z^2 I_{\bar{d}})$, where κ_z^2 is chosen by users. We adjust the value of κ_z^2 in the proposal distribution, such that the acceptance rate of $\tilde{z}_{i,g}$ is between 20% and 40%.
Let $\widetilde{Z}_g = (z_{1,g}^{(t-1)}, \cdots, z_{i-1,g}^{(t-1)}, \tilde{z}_{i,g}, z_{i+1,g}^{(t-1)}, \cdots, z_{m_g,g}^{(t-1)})$.

(b) With the probability equal to

$$\alpha(z_{i,g}^{(t-1)}; \tilde{z}_{i,g}) =$$
$$\min\left\{ \frac{P(Y_g|W_g, \widetilde{Z}_g; \lambda^{(t-1)}, \beta^{(t-1)}, \sigma^{(t-1)}, \alpha_g^{(t-1)})}{P(Y_g|W_g, Z_g^{(t-1)}; \lambda^{(t-1)}, \beta^{(t-1)}, \sigma^{(t-1)}, \alpha_g^{(t-1)})} \cdot \frac{\mathcal{N}_{\bar{d}}(\tilde{z}_{i,g}; 0, I_{\bar{d}})}{\mathcal{N}_{\bar{d}}(z_{i,g}^{(t-1)}; 0, I_{\bar{d}})}, 1 \right\},$$

set $z_{i,g}^{(t)}$ equal to $\tilde{z}_{i,g}$. Otherwise, set it to $z_{i,g}^{(t-1)}$.

Step 2. Simulate $\phi^{(t)} = (\gamma^{(t)'}, \delta^{(t)'})$ from $P(\phi|\{W_g\}, \{Z_g^{(t)}\})$, $\phi \in O_1$, by using the M-H algorithm.

(a) Propose $\tilde{\phi} \sim \mathcal{N}_{2\bar{s}+\bar{q}+\bar{d}}\left(\phi^{(t-1)}, \kappa_\phi^2 I_{2\bar{s}+\bar{q}+\bar{d}}\right)$, where κ_ϕ^2 is chosen by users.

(b) With the probability equal to

$$\alpha(\phi^{(t-1)}; \tilde{\phi}) =$$
$$\min\left\{ \prod_{g=1}^{G} \left(\frac{P(W_g|Z_g^{(t)}, \tilde{\phi})}{P(W_g|Z_g^{(t)}, \phi^{(t-1)})} \right) \cdot \frac{\mathcal{N}_{2\bar{s}+\bar{q}+\bar{d}}(\tilde{\phi}; \phi_0, \Phi_0)}{\mathcal{N}_{2\bar{s}+\bar{q}+\bar{d}}(\phi^{(t-1)}; \phi_0, \Phi_0)} \cdot \frac{I(\tilde{\phi} \in O_1)}{I(\phi^{(t-1)} \in O_1)}, 1 \right\}$$

set $\phi^{(t)}$ equal to $\tilde{\phi}$. Otherwise, set it to $\phi^{(t-1)}$.

Step 3. Sample $\lambda^{(t)}$ from $P(\lambda|\{Y_g\},\{W_g\},\{Z_g^{(t)}\},\beta^{(t-1)},\sigma^{(t-1)},\{\alpha_g^{(t-1)}\})$, $\lambda\in[-1/\tau_G,1/\tau_G]$, as specified in Eq. (3), by using the M-H algorithm.

(a) Propose $\tilde{\lambda}\sim\mathscr{N}\left(\lambda^{(t-1)},\kappa_\lambda^2\right)$, where κ_λ^2 is chosen by users.

(b) Let $A=[-1/\tau_G,1/\tau_G]$, with the probability equal to

$$\alpha(\lambda^{(t-1)};\tilde{\lambda})=$$
$$\min\left\{\prod_{g=1}^{G}\left(\frac{p(Y_g|W_g,Z_g^{(t)};\tilde{\lambda},\beta^{(t-1)},\sigma^{(t-1)},\alpha_g^{(t-1)})}{p(Y_g|W_g,Z_g^{(t)};\lambda^{(t-1)},\beta^{(t-1)},\sigma^{(t-1)},\alpha_g^{(t-1)})}\right)\cdot\frac{I(\tilde{\lambda}\in A)}{I(\lambda^{(t-1)}\in A)},1\right\},$$

set $\lambda^{(t)}$ equal to $\tilde{\lambda}$. Otherwise, set it to $\lambda^{(t-1)}$.

Step 4. Sample $\beta^{(t)}$ from $P(\beta|\{Y_g\},\{W_g\},\{Z_g^{(t)}\},\lambda^{(t)},\sigma^{(t-1)},\{\alpha_g^{(t-1)}\})$, as specified in Eq. (4).

Step 5. Sample $\sigma^{(t)}$ from $P(\sigma|\{Y_g\},\{W_g\},\{Z_g^{(t)}\},\lambda^{(t)},\beta^{(t)},\{\alpha_g^{(t-1)}\})$, $\sigma\in O_2$, as specified in Eq. (5), by using the M-H algorithm.

(a) Propose $\tilde{\sigma}\sim\mathscr{N}_{\bar{d}+1}\left(\sigma^{(t-1)},\kappa_\sigma^2 I_{\bar{d}+1}\right)$, where κ_σ^2 is chosen by users.

(b) with the probability

$$\alpha(\sigma^{(t-1)};\tilde{\sigma})=\min\left\{\prod_{g=1}^{G}\left(\frac{p(Y_g|W_g,Z_g^{(t)};\lambda^{(t)},\beta^{(t)},\tilde{\sigma},\alpha_g^{(t-1)})}{p(Y_g|W_g,Z_g^{(t)};\lambda^{(t)},\beta^{(t)},\sigma^{(t-1)},\alpha_g^{(t-1)})}\right)\cdot\frac{\mathscr{N}_{\bar{d}+1}(\tilde{\sigma}|\sigma_0,\Sigma_0)}{\mathscr{N}_{\bar{d}+1}(\sigma^{(t-1)}|\sigma_0,\Sigma_0)}\cdot\frac{I(\tilde{\sigma}\in O_2)}{I(\sigma^{(t-1)}\in O_2)},1\right\}$$

set $\sigma^{(t)}$ equal to $\tilde{\sigma}$. Otherwise, set it to $\sigma^{(t-1)}$.

Step 6. Sample $\alpha_g^{(t)}$ from $P(\alpha_g|Y_g,W_g,Z_g^{(t)},\lambda^{(t)},\beta^{(t)},\sigma^{(t)})$, as specified in Eq. (6) for $g=1,\cdots,G$.

Appendix II. Diagnostics of the MCMC convergence

An unexpected slow convergence of the MCMC sampling could be due to either programming mistakes, an inadequate model that does not fit the data, or simply the slow movement of the Markov chain. Therefore, making sure that the statistical inference is based on the converged MCMC results (Gelman 1996) is crucial. One way to evaluate the convergence of the MCMC sampling is visual inspection. More formally, we consider methods proposed by Geweke (1992), Raftery and Lewis (1992), and Heidelberger and Welch (1983). Geweke (1992) compares the sample means of the first $x\%$ of the MCMC draws versus the last $y\%$ of the MCMC draws. The null hypothesis of an equal mean is examined by using the standard Z-score test, with standard errors estimated numerically. Raftery and Lewis (1992) study the quantiles of the posterior distribution for the parameter vector θ. Suppose one wants to estimate the posterior probability, $P(\theta \leq u|\text{data})$, to within a range of error $(-r, +r)$ with a probability s, where u is a particular cutoff corresponding to a specified quantile. Raftery and Lewis propose a method to calculate the (approximate) required number of MCMC iterations when the actual quantile of interest is q. The idea of Heidelberger-Welch diagnostic is to use the Cramér-von Mises statistic to test the null hypothesis of stationarity.

All three methods are available in the package CODA (Best *et al.*, 1996). which supports the software such as Fortran, S-Plus, R, and Matlab. To implement Geweke's diagnostic, we provide the MCMC draws as inputs and specify values of two percentage points, $x = 50$ and $y = 50$. To implement Raftery and Lewis's diagnostic, we specify $q = 0.025$, $r = 0.005$, and $s = 0.95$, which require the cumulative distribution function of the 2.5% quantile to be estimated within the error range of ± 0.005 with the probability of 0.95. In our study, we utilize three methods as guidance to decide the length of the MCMC sampling. First we apply Raftery and Lewis's method to check the adequate sampling length. In any case, if the required sampling length is too long to be computationally feasible, we will further check

Geweke's and Heidelberger-Welch's diagnostics and finish the sampling whenever the tests are passed.

Appendix III. Bayesian analysis for missing observations

Let Y_g^o ($n_g \times 1$) denote the observed dependent variables, and Y_g^m ($(m_g - n_g) \times 1$) denote the missing observations in group g. We have

$$Y_g | W_g, X_g, Z_g, \theta = \left. \begin{pmatrix} Y_g^o \\ Y_g^m \end{pmatrix} \right| W_g, X_g, Z_g, \theta \sim \mathscr{N}_{m_g}(\mu_g, \Sigma_g),$$

where $\mu_g = \begin{pmatrix} \mu_g^o \\ \mu_g^m \end{pmatrix}$ and $\Sigma_g = \begin{pmatrix} \Sigma_g^{oo} & \Sigma_g^{om} \\ \Sigma_g^{mo} & \Sigma_g^{mm} \end{pmatrix}$. The conditional distribution of Y_g^m, given Y_g^o and others, can be written as

$$Y_g^m | Y_g^o, W_g, X_g, Z_g, \theta \sim \mathscr{N}_{m_g - n_g} \left(\mathrm{E}(Y_g^m | Y_g^o, W_g, X_g, Z_g, \theta), \mathrm{V}(Y_g^m | Y_g^o, W_g, X_g, Z_g, \theta) \right),$$

where

$$\begin{aligned} \mathrm{E}(Y_g^m | Y_g^o, W_g, X_g, Z_g, \theta) &= \mu_g^m + \Sigma_g^{mo} (\Sigma_g^{oo})^{-1} (Y_g^o - \mu_g^o), \\ \mathrm{V}(Y_g^m | Y_g^o, W_g, X_g, Z_g, \theta) &= \Sigma_g^{mm} - \Sigma_g^{mo} (\Sigma_g^{oo})^{-1} \Sigma_g^{om}. \end{aligned}$$

To deal with the problem of missing observations in the Bayesian estimation, we consider the approach of data augmentation. In the imputation step, we draw Y_g^m from the density $f(Y_g^m | Y_g^o, W_g, X_g, Z_g, \theta)$. In the posterior step, we draw θ from the density $f(\theta | Y_g^o, Y_g^m, W_g, X_g, Z_g)$. Under the SCSAR model, we have $\mu_g = S_g^{-1}(\mathbf{X_g}\beta + Z_g \sigma_{z\epsilon} + l_g \alpha_g)$ and $\Sigma_g = S_g^{-1} \left((\sigma_\epsilon^2 - \sigma_{\epsilon z} \sigma_{z\epsilon}) I_{m_g} \right) S_g^{-1'}$. Hence,

$$Y_{m,g} \sim \mathscr{N}_{m_g - n_g} \left((S_g^{-1}(\mathbf{X_g}\beta + Z_g \sigma_{z\epsilon} + l_g \alpha_g))_m + \Sigma_g^{mo} (\Sigma_g^{oo})^{-1} (Y_g^o - \mu_g^o), \Sigma_{mm}^g - \Sigma_{mo}^g \Sigma_{oo}^{g-1} \Sigma_{om}^g \right).$$

Appendix IV: Derivation of the AICM

The conventional AIC (Akaike 1973) is defined as

$$\text{AIC} = 2d - 2\ell_{\max}, \tag{7}$$

where $\ell_{\max}$ is the maximum log-likelihood and d is the dimension of the parameters in the model. $\ell_{\max}$ is not directly observable in the Bayesian estimation approach because $\ell_{\max}$ may not be reached during the MCMC sampling procedure; however, following Raftery *et al.* (2007), it may be estimated given the posterior distribution of the log-likelihoods,

$$\ell_{\max} - \ell_t \sim \text{Gamma}(d/2, 1), \tag{8}$$

where $\{\ell_t : t = 1, \cdots, T\}$ is a sequence of log-likelihoods from MCMC posterior draws with a proper thinning such that they are approximately independent. The distributional assumption in Eq. (8) is asymptotically evident when the amount of data underlying the likelihoods increases to infinity (Bickel and Ghosh 1990; Dawid 1991). Based on the Gamma distribution, we know $E[\ell_{\max} - \ell_t] = d/2$ and $\text{Var}(\ell_t) = d/2$. Therefore, we can obtain the moment estimators $\hat{d} = 2s_\ell^2$ and $\hat{\ell}_{\max} = \bar{\ell} + s_\ell^2$, where $\bar{\ell}$ and s_ℓ^2 are the sample mean and variance of the ℓ_t's, respectively. The simulation-based (Monte Carlo) version of AIC is given as

$$\text{AICM} = 2\hat{d} - 2\hat{\ell}_{\max} = 2(s_\ell^2 - \bar{\ell}). \tag{9}$$

and its standard error can be calculated by

$$\text{S.E.}(\text{AICM}) = \sqrt{4\hat{d}/(2T) + 4\hat{d}(11\hat{d}/4 + 12)/T} \tag{10}$$

by using the fact that $\text{Var}(\bar{\ell}) \approx d/(2T)$ and $\text{Var}(s_\ell^2) \approx d(11d/4 + 12)/T$ and the approximate independence between $\bar{\ell}$ and s_ℓ^2.

References

[1] Akaike H. 1973. Information theory and an extension of the maximum likelihood principle. In B. N. Petrov and F. Csaki (Eds.), Second international symposium on information theory, 267-281. Budapest: Academiai Kiado.

[2] Best N., Cowles M. K., Vines K. 1996. CODA: convergence diagnosis and output analysis software for gibbs sampling output version 0.30.

[3] Bickel P. J., Ghosh J. K. 1990. Decomposition of the likelihood ratio statistic and the Bartlett correction – A Bayesian argument. *Annals of Statistics*, 18, 1070-1090.

[4] Chib S., Greenberg E. 1996. Markov chain Monte Carlo simulation methods in econometrics. *Econometric Theory* 12: 409-431.

[5] Dawid A. P. 1991. Fisherian inference in likelihood and prequential frames of reference. *Journal of Royal Statistical Society B*, 53, 79-109, (with discussion).

[6] Gelman A. 1996. Inference and monitoring convergence. In Gilks, W. R., S. Richardson and D. J. Spiegelhalter (Eds.), *Markov Chain Monte Carlo in Practice.*

[7] Geweke J. 1992. Evaluating the accuracy of sampling-based approaches to calculating posterior moments. *Bayesian Statistics* 4, (ed. J. M. Bernardo, J. O. Berger, A. P. Dawid, and A. F. M. Smith). Clarendon Press, Oxford, UK.

[8] Handcock M. S., Raftery A. E., Tantrum J. M. 2007. Model-based clustering for social networks. *Journal of the Royal Sataistical Society* 170, 301-54.

[9] Heidelberger P., Welch P.D. 1983. Simulation Run Length Control in the Presence of an Initial Transient. *Operations Research*, 31, p. 1109 1144.

[10] Raftery A. E., Lewis S. M. 1992. How many iterations in the Gibbs sampler? *Bayesian Statistics* 4.

[11] Raftery A. E., Newton M. A., Satagopan J. M., Krivitsky P. N. 2007. Estimating the integrated likelihood via posterior simulation using the harmonic mean identity. *Bayesian Statistics*, 8, 1-45.

[12] Tierney L. 1994. Markov chains for exploring posterior distributions (with discussion). *Annuals of Statistics* 22: 1701-1062.