A Count Data Model with Social Interactions

A. Houndetoungan

Department of Economics
Université Laval

February 3, 2021

Social Network Models

- **Why is it important to estimate peer effects? (Manski 1993, REStud)**
  - E.g, Participation in extracurricular activities.
  - Decrease in the number of hours in class; Student increases his participation; Student's friends increase their participation;
  - Because Student's friends increase their participation, Student further increases his participation; ...
  - Social multiplier increasing the impact of exogenous shocks (direct impact due exogenous shocks + indirect impact because friends change their behavior).

- Example of model

$$\text{Behavior} = F\left(\text{Friend's Behavior}, \text{Control Variables}\right)$$

  - Peer effects in adolescent overweight (Trogdon, Nonnemaker, and Pais 2008, JHE);
  - Peer effects in education (Calvó-Armengol, Patacchini, and Zenou 2009, REStud);
  - Peer effects in the workplace (Cornelissen, Dustmann, and Schönberg 2017, AER).

Social Network Models

- **Why is it important to estimate peer effects? (Manski 1993, REStud)**
  - E.g, Participation in extracurricular activities.
  - Decrease in the number of hours in class; Student increases his participation; Student's friends increase their participation;
  - Because Student's friends increase their participation, Student further increases his participation; ...
  - Social multiplier increasing the impact of exogenous shocks (direct impact due exogenous shocks + indirect impact because friends change their behavior).

- Example of model

$$\text{Behavior} = F\,(\text{Friend's Behavior}, \text{Control Variables})$$

  - Peer effects in adolescent overweight (Trogdon, Nonnemaker, and Pais 2008, JHE);
  - Peer effects in education (Calvó-Armengol, Patacchini, and Zenou 2009, REStud);
  - Peer effects in the workplace (Cornelissen, Dustmann, and Schönberg 2017, AER).

Social Network Models

- **Why is it important to estimate peer effects? (Manski 1993, REStud)**
  - E.g, Participation in extracurricular activities.
  - Decrease in the number of hours in class; Student increases his participation; Student's friends increase their participation;
  - Because Student's friends increase their participation, Student further increases his participation; ...
  - Social multiplier increasing the impact of exogenous shocks (direct impact due exogenous shocks + indirect impact because friends change their behavior).

- Example of model

$$\text{Behavior} = F\left(\text{Friend's Behavior}, \text{Control Variables}\right)$$

  - Peer effects in adolescent overweight (Trogdon, Nonnemaker, and Pais 2008, JHE);
  - Peer effects in education (Calvó-Armengol, Patacchini, and Zenou 2009, REStud);
  - Peer effects in the workplace (Cornelissen, Dustmann, and Schönberg 2017, AER).

Social Network Models

- **Why is it important to estimate peer effects? (Manski 1993, REStud)**
  - E.g, Participation in extracurricular activities.
  - Decrease in the number of hours in class; Student increases his participation; Student's friends increase their participation;
  - Because Student's friends increase their participation, Student further increases his participation; ...
  - Social multiplier increasing the impact of exogenous shocks (direct impact due exogenous shocks + indirect impact because friends change their behavior).

- Example of model

$$\text{Behavior} = F\left(\text{Friend's Behavior}, \text{Control Variables}\right)$$

  - Peer effects in adolescent overweight (Trogdon, Nonnemaker, and Pais 2008, JHE);
  - Peer effects in education (Calvó-Armengol, Patacchini, and Zenou 2009, REStud);
  - Peer effects in the workplace (Cornelissen, Dustmann, and Schönberg 2017, AER).

Social Network Models

- **Why is it important to estimate peer effects? (Manski 1993, REStud)**
  - E.g, Participation in extracurricular activities.
  - Decrease in the number of hours in class; Student increases his participation; Student's friends increase their participation;
  - Because Student's friends increase their participation, Student further increases his participation; . . .
  - Social multiplier increasing the impact of exogenous shocks (direct impact due exogenous shocks + indirect impact because friends change their behavior).

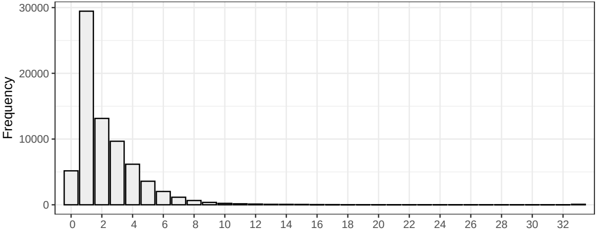
- Example of model

$$\text{Behavior} = F\left(\text{Friend's Behavior}, \text{Control Variables}\right)$$

  - Peer effects in adolescent overweight (Trogdon, Nonnemaker, and Pais 2008, JHE);
  - Peer effects in education (Calvó-Armengol, Patacchini, and Zenou 2009, REStud);
  - Peer effects in the workplace (Cornelissen, Dustmann, and Schönberg 2017, AER).

Social Network Models and Discrete Data

- Example of count data from Add Health: Number of extracurricular activities in which students are enrolled.

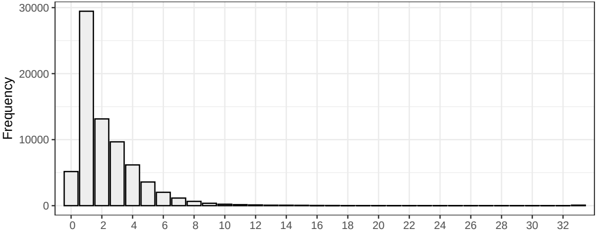


Integer values and left censored at zero

- Models with social interactions:
  1. Linear-in-means model (Bramoullé, Djebbari, and Fortin 2009, JE), (L.-F. Lee 2004, Econometrica);
  2. Binary data (Brock and Durlauf 2001, REStud), (Brock and Durlauf 2001, REStat);
  3. Censored choices (Xu and L.-f. Lee 2015, JE).

Social Network Models and Discrete Data

- Example of count data from Add Health: Number of extracurricular activities in which students are enrolled.

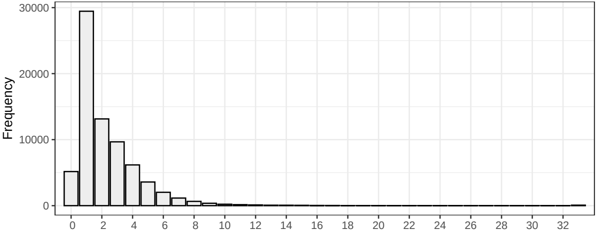


Integer values and left censored at zero

- Models with social interactions:
  1. Linear-in-means model (Bramoullé, Djebbari, and Fortin 2009, JE), (L.-F. Lee 2004, Econometrica);
  2. Binary data (Brock and Durlauf 2001, REStud), (Brock and Durlauf 2001, REStat);
  3. Censored choices (Xu and L.-f. Lee 2015, JE).

Social Network Models and Discrete Data

- Example of count data from Add Health: Number of extracurricular activities in which students are enrolled.

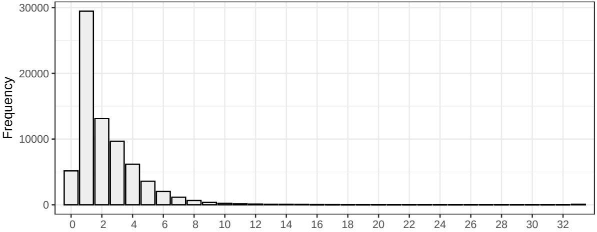


Integer values and left censored at zero

- Models with social interactions:
  1. Linear-in-means model (Bramoullé, Djebbari, and Fortin 2009, JE), (L.-F. Lee 2004, Econometrica);
  2. Binary data (Brock and Durlauf 2001, REStud), (Brock and Durlauf 2001, REStat);
  3. Censored choices (Xu and L.-f. Lee 2015, JE).

Social Network Models and Discrete Data

- Example of count data from Add Health: Number of extracurricular activities in which students are enrolled.



Integer values and left censored at zero

- Models with social interactions:
  1. Linear-in-means model (Bramoullé, Djebbari, and Fortin 2009, JE), (L.-F. Lee 2004, Econometrica);
  2. Binary data (Brock and Durlauf 2001, REStud), (Brock and Durlauf 2001, REStat);
  3. Censored choices (Xu and L.-f. Lee 2015, JE).

This paper

1. Model of random utility dealing with networks and count choices.
   - Number of count choices is unbounded;
   - Game of incomplete information.
2. Generalization of Rational Expectation model presented by L.-f. Lee, Li, and Lin 2014 (REStat) for binary outcome.
3. (Under some conditions, e.g, when the number of count choices is large) my model is asymptotically similar to the linear models;
   - Linear Spatial Autoregressive (SAR) model (L.-F. Lee 2004, Econometrica);
   - Spatial Autoregressive Tobit (SART) model (Xu and L.-f. Lee 2015, JE).

This paper

1. Model of random utility dealing with networks and count choices.
   - Number of count choices is unbounded;
   - Game of incomplete information.
2. Generalization of Rational Expectation model presented by L.-f. Lee, Li, and Lin 2014 (REStat) for binary outcome.
3. (Under some conditions, e.g, when the number of count choices is large) my model is asymptotically similar to the linear models;
   - Linear Spatial Autoregressive (SAR) model (L.-F. Lee 2004, Econometrica);
   - Spatial Autoregressive Tobit (SART) model (Xu and L.-f. Lee 2015, JE).

This paper

1. Model of random utility dealing with networks and count choices.
   - Number of count choices is unbounded;
   - Game of incomplete information.

2. Generalization of Rational Expectation model presented by L.-f. Lee, Li, and Lin 2014 (REStat) for binary outcome.

3. (Under some conditions, e.g, when the number of count choices is large) my model is asymptotically similar to the linear models;
   - Linear Spatial Autoregressive (SAR) model (L.-F. Lee 2004, Econometrica);
   - Spatial Autoregressive Tobit (SART) model (Xu and L.-f. Lee 2015, JE).

This paper

1. Model of random utility dealing with networks and count choices.
   - Number of count choices is unbounded;
   - Game of incomplete information.

2. Generalization of Rational Expectation model presented by L.-f. Lee, Li, and Lin 2014 (REStat) for binary outcome.

3. (Under some conditions, e.g, when the number of count choices is large) my model is asymptotically similar to the linear models;
   - Linear Spatial Autoregressive (SAR) model (L.-F. Lee 2004, Econometrica);
   - Spatial Autoregressive Tobit (SART) model (Xu and L.-f. Lee 2015, JE).

Empirical Results

1. (**Application**) Peer effects on the number of extracurricular activities in which students are enrolled.
   - **Peer marginal effect:** 0.294;
   - SART model: 0.141, SAR model 0.166;
2. Endogeneity of the network controlled.
   - Unobserved variables such as sociability degree may explain the network and the participation in extracurricular activities;
   - Do not take into account the endogeneity of the network significantly overestimates the peer effects.
3. An easy to use R package—named `CDatanet`— located on my GitHub implementing the model.

Empirical Results

1. (**Application**) Peer effects on the number of extracurricular activities in which students are enrolled.
   - **Peer marginal effect:** 0.294;
   - SART model: 0.141, SAR model 0.166;
2. Endogeneity of the network controlled.
   - Unobserved variables such as sociability degree may explain the network and the participation in extracurricular activities;
   - Do not take into account the endogeneity of the network significantly overestimates the peer effects.
3. An easy to use R package—named `CDatanet`— located on my GitHub implementing the model.

Empirical Results

1. (**Application**) Peer effects on the number of extracurricular activities in which students are enrolled.
   - **Peer marginal effect:** 0.294;
   - SART model: 0.141, SAR model 0.166;
2. Endogeneity of the network controlled.
   - Unobserved variables such as sociability degree may explain the network and the participation in extracurricular activities;
   - Do not take into account the endogeneity of the network significantly overestimates the peer effects.
3. An easy to use R package—named `CDatanet`— located on my GitHub implementing the model.

Empirical Results

1. (**Application**) Peer effects on the number of extracurricular activities in which students are enrolled.
   - **Peer marginal effect:** 0.294;
   - SART model: 0.141, SAR model 0.166;

2. Endogeneity of the network controlled.
   - Unobserved variables such as sociability degree may explain the network and the participation in extracurricular activities;
   - Do not take into account the endogeneity of the network significantly overestimates the peer effects.

3. An easy to use R package—named `CDatanet`— located on my GitHub implementing the model.

Empirical Results

1. (**Application**) Peer effects on the number of extracurricular activities in which students are enrolled.
   - **Peer marginal effect:** 0.294;
   - SART model: 0.141, SAR model 0.166;
2. Endogeneity of the network controlled.
   - Unobserved variables such as sociability degree may explain the network and the participation in extracurricular activities;
   - Do not take into account the endogeneity of the network significantly overestimates the peer effects.
3. An easy to use R package—named `CDatanet`— located on my GitHub implementing the model.

Empirical Results

1. (**Application**) Peer effects on the number of extracurricular activities in which students are enrolled.
   - **Peer marginal effect:** 0.294;
   - SART model: 0.141, SAR model 0.166;
2. Endogeneity of the network controlled.
   - Unobserved variables such as sociability degree may explain the network and the participation in extracurricular activities;
   - Do not take into account the endogeneity of the network significantly overestimates the peer effects.
3. An easy to use R package—named `CDatanet`— located on my GitHub implementing the model.

Empirical Results

1. (**Application**) Peer effects on the number of extracurricular activities in which students are enrolled.
   - **Peer marginal effect:** 0.294;
   - SART model: 0.141, SAR model 0.166;
2. Endogeneity of the network controlled.
   - Unobserved variables such as sociability degree may explain the network and the participation in extracurricular activities;
   - Do not take into account the endogeneity of the network significantly overestimates the peer effects.
3. An easy to use R package—named `CDatanet`— located on my GitHub implementing the model.

Outline

Game: Main Assumption

- Individuals choose a continuous latent variable y_i^* (interpreted as an intention, see Maddala 1986) which determines y_i (the observed variable).

- Binary choices (L.-f. Lee, Li, and Lin 2014; Liu 2019).

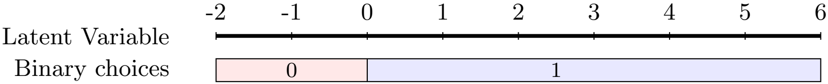
- Assumption for count variable (see Cameron and Trivedi 1990).

- (**Generalization of binary models**) If $\gamma = \infty$, $y_i \in \{0, 1\}$.

Game: Main Assumption

- Individuals choose a continuous latent variable y_i^* (interpreted as an intention, see Maddala 1986) which determines y_i (the observed variable).

- Binary choices (L.-f. Lee, Li, and Lin 2014; Liu 2019).



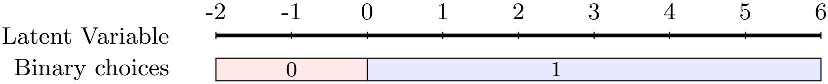
- Assumption for count variable (see Cameron and Trivedi 1990).

- (**Generalization of binary models**) If $\gamma = \infty$, $y_i \in \{0, 1\}$.

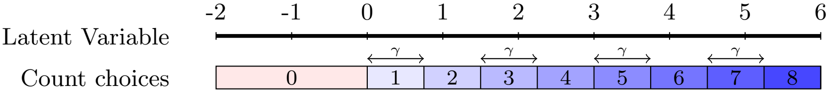
Game: Main Assumption

- Individuals choose a continuous latent variable y_i^* (interpreted as an intention, see Maddala 1986) which determines y_i (the observed variable).

- Binary choices (L.-f. Lee, Li, and Lin 2014; Liu 2019).



- Assumption for count variable (see Cameron and Trivedi 1990).

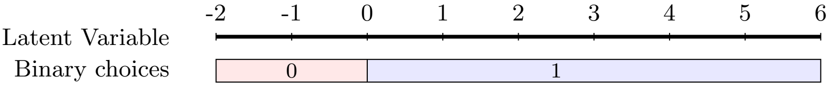


- (**Generalization of binary models**) If $\gamma = \infty$, $y_i \in \{0, 1\}$.

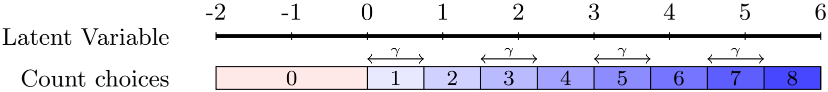
Game: Main Assumption

- Individuals choose a continuous latent variable y_i^* (interpreted as an intention, see Maddala 1986) which determines y_i (the observed variable).

- Binary choices (L.-f. Lee, Li, and Lin 2014; Liu 2019).



- Assumption for count variable (see Cameron and Trivedi 1990).



- (**Generalization of binary models**) If $\gamma = \infty$, $y_i \in \{0, 1\}$.

Game: Preferences

- Preferences (see also Ballester, Calvó-Armengol, and Zenou 2006; Calvó-Armengol, Patacchini, and Zenou 2009).

$$\mathcal{U}_i = \underbrace{(\psi_i + \varepsilon_i)\, y_i^* - \frac{y_i^{*2}}{2}}_{\text{private sub-utility}} + \underbrace{\lambda y_i^* \sum_{j \neq i} g_{ij} y_j}_{\text{social sub-utility}} \tag{1}$$

where ψ_i, $\lambda \in \mathbb{R}$ and ε_i is a private information with a common distribution known among individuals.

- Expected utility

$$\mathbf{E}\left(\mathcal{U}_i | y_i^*, \varepsilon_i, \lambda, \boldsymbol{\psi}, \mathbf{G}\right) = (\psi_i + \varepsilon_i)\, y_i^* - \frac{y_i^{*2}}{2} + \lambda y_i^* \sum_{j \neq i} g_{ij} \bar{y}_j, \tag{2}$$

where $\forall\ j \in \mathcal{V}$,

$$\bar{y}_j = \sum_{r=0}^{\infty} r p_{jr} \tag{3}$$

and p_{jr} is the probability of $y_j = r$.

Game: Preferences

- Preferences (see also Ballester, Calvó-Armengol, and Zenou 2006; Calvó-Armengol, Patacchini, and Zenou 2009).

$$\mathcal{U}_i = \underbrace{(\psi_i + \varepsilon_i)\, y_i^* - \frac{y_i^{*2}}{2}}_{\text{private sub-utility}} + \underbrace{\lambda y_i^* \sum_{j \neq i} g_{ij} y_j}_{\text{social sub-utility}} \tag{1}$$

where ψ_i, $\lambda \in \mathbb{R}$ and ε_i is a private information with a common distribution known among individuals.

- Expected utility

$$\mathbf{E}\left(\mathcal{U}_i | y_i^*, \varepsilon_i, \lambda, \boldsymbol{\psi}, \mathbf{G}\right) = (\psi_i + \varepsilon_i)\, y_i^* - \frac{y_i^{*2}}{2} + \lambda y_i^* \sum_{j \neq i} g_{ij} \bar{y}_j, \tag{2}$$

where $\forall\ j \in \mathcal{V}$,

$$\bar{y}_j = \sum_{r=0}^{\infty} r p_{jr} \tag{3}$$

and p_{jr} is the probability of $y_j = r$.

Game: First Order Conditions (focs)

- focs

$$y_i^* = \lambda \mathbf{g}_i \bar{\mathbf{y}} + \mathbf{x}_i' \boldsymbol{\beta} + \varepsilon_i, \tag{4}$$

- Belief at equilibrium for all $i = 1, \dots, n$ and $q \in \mathbb{N}$

$$p_{iq} = \mathcal{P}\left(y_i^* \in (a_{q-1},\ a_q)\right)$$
$$p_{iq} = \mathrm{F}_\varepsilon\left(\lambda \mathbf{g}_i \bar{\mathbf{y}} + \mathbf{x}_i' \boldsymbol{\beta} - a_q\right) - \mathrm{F}_\varepsilon\left(\lambda \mathbf{g}_i \bar{\mathbf{y}} + \mathbf{x}_i' \boldsymbol{\beta} - a_{q+1}\right) \tag{5}$$

- $\bar{y}_i = \sum_{r=0}^{\infty} r p_{ir}$. $\implies$ Bijective function between (p_{iq}) and $(\bar{y}_i)$.

- Fixed point equation: $\bar{y}_i = \mathbf{L}(\bar{\mathbf{y}})$.

VS Poisson

$$\bar{y}_i = \sum_{r=1}^{\infty} \mathrm{F}_\varepsilon\left(\lambda \mathbf{g}_i \bar{\mathbf{y}} + \mathbf{x}_i' \boldsymbol{\beta} - a_r\right) \tag{6}$$

Game: First Order Conditions (focs)

- focs

$$y_i^* = \lambda \mathbf{g}_i \bar{\mathbf{y}} + \mathbf{x}_i' \boldsymbol{\beta} + \varepsilon_i, \tag{4}$$

- Belief at equilibrium for all $i = 1, \dots, n$ and $q \in \mathbb{N}$

$$p_{iq} = \mathcal{P}\left(y_i^* \in (a_{q-1},\ a_q)\right)$$
$$p_{iq} = \mathrm{F}_\varepsilon\left(\lambda \mathbf{g}_i \bar{\mathbf{y}} + \mathbf{x}_i' \boldsymbol{\beta} - a_q\right) - \mathrm{F}_\varepsilon\left(\lambda \mathbf{g}_i \bar{\mathbf{y}} + \mathbf{x}_i' \boldsymbol{\beta} - a_{q+1}\right) \tag{5}$$

- $\bar{y}_i = \sum_{r=0}^{\infty} r p_{ir}$. $\implies$ Bijective function between (p_{iq}) and $(\bar{y}_i)$.

- Fixed point equation: $\bar{y}_i = \mathbf{L}(\bar{\mathbf{y}})$.

VS Poisson

$$\bar{y}_i = \sum_{r=1}^{\infty} \mathrm{F}_\varepsilon\left(\lambda \mathbf{g}_i \bar{\mathbf{y}} + \mathbf{x}_i' \boldsymbol{\beta} - a_r\right) \tag{6}$$

- focs

$$y_i^* = \lambda \mathbf{g}_i \bar{\mathbf{y}} + \mathbf{x}_i' \boldsymbol{\beta} + \varepsilon_i, \tag{4}$$

- Belief at equilibrium for all $i = 1, \dots, n$ and $q \in \mathbb{N}$

$$\begin{aligned} p_{iq} &= \mathcal{P}\left(y_i^* \in (a_{q-1},\ a_q)\right) \\ p_{iq} &= \mathrm{F}_\varepsilon\left(\lambda \mathbf{g}_i \bar{\mathbf{y}} + \mathbf{x}_i' \boldsymbol{\beta} - a_q\right) - \mathrm{F}_\varepsilon\left(\lambda \mathbf{g}_i \bar{\mathbf{y}} + \mathbf{x}_i' \boldsymbol{\beta} - a_{q+1}\right) \end{aligned} \tag{5}$$

- $\bar{y}_i = \sum_{r=0}^{\infty} r p_{ir}$. $\implies$ Bijective function between (p_{iq}) and $(\bar{y}_i)$.

- Fixed point equation: $\bar{y}_i = \mathbf{L}(\bar{\mathbf{y}})$. VS Poisson

$$\bar{y}_i = \sum_{r=1}^{\infty} \mathrm{F}_\varepsilon\left(\lambda \mathbf{g}_i \bar{\mathbf{y}} + \mathbf{x}_i' \boldsymbol{\beta} - a_r\right) \tag{6}$$

Game: First Order Conditions (focs)

- focs

$$y_i^* = \lambda \mathbf{g}_i \bar{\mathbf{y}} + \mathbf{x}_i' \boldsymbol{\beta} + \varepsilon_i, \tag{4}$$

- Belief at equilibrium for all $i = 1, \dots, n$ and $q \in \mathbb{N}$

$$p_{iq} = \mathcal{P}\left(y_i^* \in (a_{q-1},\ a_q)\right)$$
$$p_{iq} = \mathrm{F}_{\varepsilon}\left(\lambda \mathbf{g}_i \bar{\mathbf{y}} + \mathbf{x}_i' \boldsymbol{\beta} - a_q\right) - \mathrm{F}_{\varepsilon}\left(\lambda \mathbf{g}_i \bar{\mathbf{y}} + \mathbf{x}_i' \boldsymbol{\beta} - a_{q+1}\right) \tag{5}$$

- $\bar{y}_i = \sum_{r=0}^{\infty} r p_{ir}. \implies$ Bijective function between (p_{iq}) and $(\bar{y}_i)$.

- Fixed point equation: $\bar{y}_i = \mathbf{L}(\bar{\mathbf{y}})$.

VS Poisson

$$\bar{y}_i = \sum_{r=1}^{\infty} \mathrm{F}_{\varepsilon}\left(\lambda \mathbf{g}_i \bar{\mathbf{y}} + \mathbf{x}_i' \boldsymbol{\beta} - a_r\right) \tag{6}$$

Game: Equilibrium

- Equilibrium conditions
  - Distribution of ε_i is continuous, with a derivable cdf, F_ε, and a pdf, f_ε which decrease exponentially in its tails;
  - $|\lambda| < \dfrac{C_{\gamma,\sigma_\varepsilon}}{||\mathbf{G}||_\infty}$, where $C_{\gamma,\sigma_\varepsilon} = \dfrac{\sigma_\varepsilon}{\max_{u\in\mathbb{R}} \sum_{k=-\infty}^{\infty} f_\varepsilon\left(\dfrac{u+\gamma k}{\sigma_\varepsilon}\right)}$.

  – If $\varepsilon_i \overset{iid}{\sim} \mathcal{N}\left(0, \sigma_\varepsilon^2\right)$,

Bounds

$$C_{\gamma,\sigma_\varepsilon} = \frac{\sigma_\varepsilon}{\phi\left(0\right) + 2\sum_{k=1}^{\infty} \phi\left(\frac{\gamma k}{\sigma_\varepsilon}\right)}$$

.

- If $\gamma = \infty$, (ii) implies $|\lambda| < \dfrac{\sigma_\varepsilon}{||\mathbf{G}||_\infty \phi(0)}$, which is the restriction set on $|\lambda|$ in binary models (L.-f. Lee, Li, and Lin 2014; Liu 2019).

- Under the equilibrium condition, $\mathbf{L}$ is a contracting mapping. The game has a unique equilibrium and there is a unique expected outcome $\bar{\mathbf{y}}$ such that $\bar{\mathbf{y}} = \mathbf{L}\left(\bar{\mathbf{y}}\right)$.

Game: Equilibrium

- Equilibrium conditions
  - Distribution of ε_i is continuous, with a derivable cdf, F_ε, and a pdf, f_ε which decrease exponentially in its tails;
  - $|\lambda| < \dfrac{C_{\gamma,\sigma_\varepsilon}}{||\mathbf{G}||_\infty}$, where $C_{\gamma,\sigma_\varepsilon} = \dfrac{\sigma_\varepsilon}{\max_{u\in\mathbb{R}} \sum_{k=-\infty}^{\infty} f_\varepsilon\left(\dfrac{u+\gamma k}{\sigma_\varepsilon}\right)}$.

  – If $\varepsilon_i \overset{iid}{\sim} \mathcal{N}\left(0, \sigma_\varepsilon^2\right)$,

$$C_{\gamma,\sigma_\varepsilon} = \frac{\sigma_\varepsilon}{\phi\left(0\right) + 2\sum_{k=1}^{\infty} \phi\left(\frac{\gamma k}{\sigma_\varepsilon}\right)}$$

.

Bounds

- If $\gamma = \infty$, (ii) implies $|\lambda| < \dfrac{\sigma_\varepsilon}{||\mathbf{G}||_\infty \phi(0)}$, which is the restriction set on $|\lambda|$ in binary models (L.-f. Lee, Li, and Lin 2014; Liu 2019).

- Under the equilibrium condition, $\mathbf{L}$ is a contracting mapping. The game has a unique equilibrium and there is a unique expected outcome $\bar{\mathbf{y}}$ such that $\bar{\mathbf{y}} = \mathbf{L}\left(\bar{\mathbf{y}}\right)$.

Game: Equilibrium

- Equilibrium conditions
  - Distribution of ε_i is continuous, with a derivable cdf, F_ε, and a pdf, f_ε which decrease exponentially in its tails;
  - $|\lambda| < \dfrac{C_{\gamma,\sigma_\varepsilon}}{||\mathbf{G}||_\infty}$, where $C_{\gamma,\sigma_\varepsilon} = \dfrac{\sigma_\varepsilon}{\max_{u\in\mathbb{R}} \sum_{k=-\infty}^{\infty} f_\varepsilon\left(\dfrac{u+\gamma k}{\sigma_\varepsilon}\right)}$.
  - If $\varepsilon_i \overset{iid}{\sim} \mathcal{N}\left(0, \sigma_\varepsilon^2\right)$,

Bounds

$$C_{\gamma,\sigma_\varepsilon} = \frac{\sigma_\varepsilon}{\phi\left(0\right) + 2\sum_{k=1}^{\infty} \phi\left(\frac{\gamma k}{\sigma_\varepsilon}\right)}$$

.

- If $\gamma = \infty$, (ii) implies $|\lambda| < \dfrac{\sigma_\varepsilon}{||\mathbf{G}||_\infty \phi(0)}$, which is the restriction set on $|\lambda|$ in binary models (L.-f. Lee, Li, and Lin 2014; Liu 2019).

- Under the equilibrium condition, $\mathbf{L}$ is a contracting mapping. The game has a unique equilibrium and there is a unique expected outcome $\bar{\mathbf{y}}$ such that $\bar{\mathbf{y}} = \mathbf{L}\left(\bar{\mathbf{y}}\right)$.

Outline

Estimation strategy

- Estimation done using the NPL algorithm proposed by Aguirregabiria and Mira 2007.

- Likelihood

$$\mathcal{L}(\theta, \bar{\mathbf{y}}) = \sum_{i=1}^{n} \sum_{r=0}^{\infty} \mathrm{I}\left\{y_i = r\right\} \log(p_{ir})$$

- Estimation
  - Start with a proposal $\bar{\mathbf{y}}_0$ for $\bar{\mathbf{y}}$;
  - Compute $\theta_1 = \arg\max_{\theta} \mathcal{L}(\theta, \bar{\mathbf{y}}_0)$ and $\mathbf{y}_1 = \mathbf{L}\left(\bar{\mathbf{y}}_0, \theta_1\right)$;
  - Compute $\theta_2 = \arg\max_{\theta} \mathcal{L}(\theta, \bar{\mathbf{y}}_1)$, $\mathbf{y}_2 = \mathbf{L}\left(\bar{\mathbf{y}}_1, \theta_2\right)$;
  - ...
  - If $\{\theta_m, \bar{\mathbf{y}}_m\}_{m\geq 1}$ converges, regardless of the initial guess $\bar{\mathbf{y}}_0$, then $\hat{\theta} = \lim_{m\to\infty} \theta_m$.

- I adapt the Proposition 2 in Aguirregabiria and Mira 2007 and prove that $\hat{\theta}$ is consistent with a normal distribution.

Estimation strategy

- Estimation done using the NPL algorithm proposed by Aguirregabiria and Mira 2007.

- Likelihood

$$\mathcal{L}(\theta, \bar{\mathbf{y}}) = \sum_{i=1}^{n} \sum_{r=0}^{\infty} \mathrm{I}\left\{y_i = r\right\} \log(p_{ir})$$

- Estimation
  - Start with a proposal $\bar{\mathbf{y}}_0$ for $\bar{\mathbf{y}}$;
  - Compute $\theta_1 = \arg\max_{\theta} \mathcal{L}(\theta, \bar{\mathbf{y}}_0)$ and $\mathbf{y}_1 = \mathbf{L}\left(\bar{\mathbf{y}}_0, \theta_1\right)$;
  - Compute $\theta_2 = \arg\max_{\theta} \mathcal{L}(\theta, \bar{\mathbf{y}}_1)$, $\mathbf{y}_2 = \mathbf{L}\left(\bar{\mathbf{y}}_1, \theta_2\right)$;
  - ...
  - If $\{\theta_m, \bar{\mathbf{y}}_m\}_{m\geq 1}$ converges, regardless of the initial guess $\bar{\mathbf{y}}_0$, then $\hat{\theta} = \lim_{m\to\infty} \theta_m$.

- I adapt the Proposition 2 in Aguirregabiria and Mira 2007 and prove that $\hat{\theta}$ is consistent with a normal distribution.

Estimation strategy

- Estimation done using the NPL algorithm proposed by Aguirregabiria and Mira 2007.
- Likelihood

$$\mathcal{L}(\boldsymbol{\theta}, \bar{\mathbf{y}}) = \sum_{i=1}^{n} \sum_{r=0}^{\infty} \mathrm{I}\left\{y_i = r\right\} \log(p_{ir})$$

- Estimation
  - Start with a proposal $\bar{\mathbf{y}}_0$ for $\bar{\mathbf{y}}$;
  - Compute $\boldsymbol{\theta}_1 = \arg\max_{\boldsymbol{\theta}} \mathcal{L}(\boldsymbol{\theta}, \bar{\mathbf{y}}_0)$ and $\mathbf{y}_1 = \mathbf{L}\left(\bar{\mathbf{y}}_0, \boldsymbol{\theta}_1\right)$;
  - Compute $\boldsymbol{\theta}_2 = \arg\max_{\boldsymbol{\theta}} \mathcal{L}(\boldsymbol{\theta}, \bar{\mathbf{y}}_1)$, $\mathbf{y}_2 = \mathbf{L}\left(\bar{\mathbf{y}}_1, \boldsymbol{\theta}_2\right)$;
  - ...
  - If $\{\boldsymbol{\theta}_m, \bar{\mathbf{y}}_m\}_{m\geq 1}$ converges, regardless of the initial guess $\bar{\mathbf{y}}_0$, then $\hat{\boldsymbol{\theta}} = \lim_{m\to\infty} \boldsymbol{\theta}_m$.
- I adapt the Proposition 2 in Aguirregabiria and Mira 2007 and prove that $\hat{\boldsymbol{\theta}}$ is consistent with a normal distribution.

Comparison with the linear model

- What happens if the econometrician estimates,

$$y_i = \tilde{\lambda}\mathbf{g}_i\mathbf{y} + \mathbf{x}_i'\tilde{\boldsymbol{\beta}} + \nu_i? \tag{7}$$

instead of the true first order condition,

$$y_i^* = \lambda \sum_{j=1}^{n} g_{ij}\bar{y}_j + \mathbf{x}_i'\boldsymbol{\beta} + \varepsilon_i \tag{8}$$

- The maximum likelihood estimator (MLE) of the parameter $\tilde{\lambda}$ based on the assumption $\nu_i \overset{iid}{\sim} \mathcal{N}\left(0, \sigma_\nu^2\right)$, where σ_ν^2 is an unknown parameter, is inconsistent.

- If $\mathbf{X}$ is a column vector of ones, the asymptotic bias of $\hat{\tilde{\lambda}}_{2SLS}$ is,

$$-\lambda \lim_{n\to\infty} \frac{\sum_{i=1}^{n} \mathbf{Var}\left(\tilde{\mathbf{g}_i}\mathbf{y}|\mathbf{X},\mathbf{G},\mathbf{Z}\right)}{\sum_{i=1}^{n} \mathbf{Var}\left(\tilde{\mathbf{g}_i}\mathbf{y}\right)} \tag{9}$$

- The bias decreases if y_i takes its values in a large range

Comparison with the linear model

- What happens if the econometrician estimates,

$$y_i = \tilde{\lambda}\mathbf{g}_i\mathbf{y} + \mathbf{x}_i'\tilde{\boldsymbol{\beta}} + \nu_i? \tag{7}$$

  instead of the true first order condition,

$$y_i^* = \lambda \sum_{j=1}^{n} g_{ij}\bar{y}_j + \mathbf{x}_i'\boldsymbol{\beta} + \varepsilon_i \tag{8}$$

- The maximum likelihood estimator (MLE) of the parameter $\tilde{\lambda}$ based on the assumption $\nu_i \overset{iid}{\sim} \mathcal{N}\left(0, \sigma_\nu^2\right)$, where σ_ν^2 is an unknown parameter, is inconsistent.

- If $\mathbf{X}$ is a column vector of ones, the asymptotic bias of $\hat{\hat{\lambda}}_{2SLS}$ is,

$$-\lambda \lim_{n\to\infty} \frac{\sum_{i=1}^{n} \mathbf{Var}\left(\tilde{\mathbf{g}_i}\mathbf{y}|\mathbf{X},\mathbf{G},\mathbf{Z}\right)}{\sum_{i=1}^{n} \mathbf{Var}\left(\tilde{\mathbf{g}_i}\mathbf{y}\right)} \tag{9}$$

- The bias decreases if y_i takes its values in a large range

Comparison with the linear model

- What happens if the econometrician estimates,

$$y_i = \tilde{\lambda}\mathbf{g}_i\mathbf{y} + \mathbf{x}_i'\tilde{\boldsymbol{\beta}} + \nu_i? \tag{7}$$

instead of the true first order condition,

$$y_i^* = \lambda \sum_{j=1}^{n} g_{ij}\bar{y}_j + \mathbf{x}_i'\boldsymbol{\beta} + \varepsilon_i \tag{8}$$

- The maximum likelihood estimator (MLE) of the parameter $\tilde{\lambda}$ based on the assumption $\nu_i \overset{iid}{\sim} \mathcal{N}\left(0, \sigma_\nu^2\right)$, where σ_ν^2 is an unknown parameter, is inconsistent.

- If $\mathbf{X}$ is a column vector of ones, the asymptotic bias of $\hat{\tilde{\lambda}}_{2SLS}$ is,

$$-\lambda \lim_{n\to\infty} \frac{\sum_{i=1}^n \mathbf{Var}\left(\tilde{\mathbf{g}_i}\mathbf{y}|\mathbf{X},\mathbf{G},\mathbf{Z}\right)}{\sum_{i=1}^n \mathbf{Var}\left(\tilde{\mathbf{g}_i}\mathbf{y}\right)} \tag{9}$$

- The bias decreases if y_i takes its values in a large range

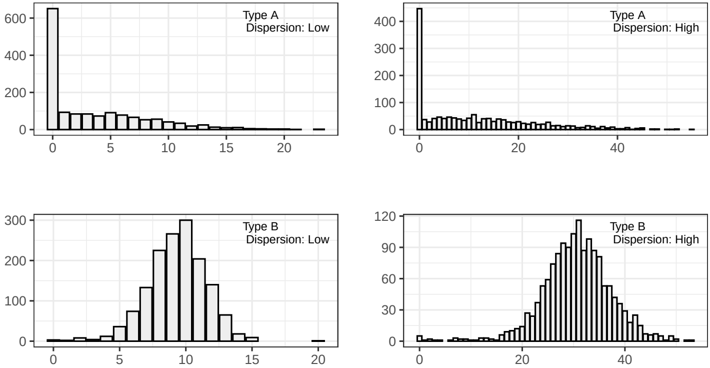
Outline

Monte Carlo simulations

- Specification

$$y_i^* = \lambda \mathbf{g}_i \bar{\mathbf{y}} + \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \gamma_1 \mathbf{g}_i \mathbf{x}_1 + \gamma_2 \mathbf{g}_i \mathbf{x}_2 + \varepsilon_i,$$

- Example of simulated data for a sample size $N = 1500$



Monte Carlo simulations

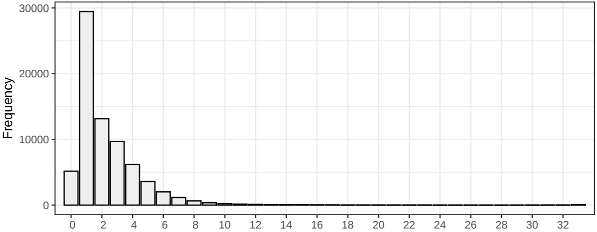
Statistic	CDSI		SART		SAR	
	Mean	Sd.	Mean	Sd.	Mean	Sd.
	Low dispersion - $N = 1500$					
	Type A					
$\lambda = 0.4$	0.402	0.088	0.268	0.078	0.143	0.132
	Type B					
$\lambda = 0.4$	0.401	0.056	0.288	0.050	0.272	0.074
	Large dispersion - $N = 1500$					
	Type A					
$\lambda = 0.4$	0.400	0.020	0.383	0.020	0.296	0.063
	Type B					
$\lambda = 0.4$	0.400	0.016	0.387	0.016	0.385	0.016

Outline

Application

- Wave I of Add Health Data: Demographic characteristics of students as well as their friendship links (i.e., best friends, up to 5 females and up to 5 males).

- Number of extracurricular activities in which students are enrolled.



- Schools with more than 100 students.

- Contextual effects and school heterogeneity as fixed effects.

Application: Exogenous network

- Network is exogenous: $\boldsymbol{\varepsilon} \perp \mathbf{G}$.

Parameters	**CDSI**			**SART**			**SAR**	
	Coef.	**Marginal Effects**		**Coef.**	**Marginal Effects**			
λ	0.443	0.363	(0.028)***	0.194	0.157	(0.005)***	0.185	(0.006)***

Application: Dyadic linking model

- Probability of link formation

$$P_{ij} = \frac{\exp\left(\Delta\mathbf{x}_{ij}'\bar{\boldsymbol{\beta}} + \mu_i + \mu_j\right)}{1 + \exp\left(\Delta\mathbf{x}_{ij}'\bar{\boldsymbol{\beta}} + \mu_i + \mu_j\right)}. \tag{10}$$

  - Observed dyad-specific variables $\Delta\mathbf{x}_{ij}$ (e.g, absolute value of age difference, indicator of same sex, ...).
  - Unobserved individual-level attribute which captures the *degree heterogeneity* μ_i (gregariousness).

- Unobserved individual-level attribute may explain y_i: $\boldsymbol{\varepsilon} \perp \mathbf{G}$ violated.

$$y_i^* = \lambda\mathbf{g}_i\bar{\mathbf{y}} + \mathbf{x}_i'\boldsymbol{\beta} + \mathbf{g}_i\mathbf{x}_i'\boldsymbol{\delta} + \overbrace{\rho\mu_i + \bar{\rho}\mathbf{g}_i\boldsymbol{\mu} + \tilde{\varepsilon}_i}^{\varepsilon_i} \tag{11}$$

- Use MCMC algorithm to estimate (10); include μ_i and $\mathbf{g}_i\boldsymbol{\mu}$ as additional explanatory variable in the count data model.

Application: Dyadic linking model

- Probability of link formation

$$P_{ij} = \frac{\exp\left(\Delta\mathbf{x}'_{ij}\bar{\boldsymbol{\beta}} + \mu_i + \mu_j\right)}{1 + \exp\left(\Delta\mathbf{x}'_{ij}\bar{\boldsymbol{\beta}} + \mu_i + \mu_j\right)}. \tag{10}$$

  – Observed dyad-specific variables $\Delta\mathbf{x}_{ij}$ (e.g, absolute value of age difference, indicator of same sex, ...).
  – Unobserved individual-level attribute which captures the *degree heterogeneity* μ_i (gregariousness).

- Unobserved individual-level attribute may explain y_i: $\boldsymbol{\varepsilon} \perp \mathbf{G}$ violated.

$$y_i^* = \lambda\mathbf{g}_i\bar{\mathbf{y}} + \mathbf{x}'_i\boldsymbol{\beta} + \mathbf{g}_i\mathbf{x}'_i\boldsymbol{\delta} + \overbrace{\rho\mu_i + \bar{\rho}\mathbf{g}_i\boldsymbol{\mu} + \tilde{\varepsilon}_i}^{\varepsilon_i} \tag{11}$$

- Use MCMC algorithm to estimate (10); include μ_i and $\mathbf{g}_i\boldsymbol{\mu}$ as additional explanatory variable in the count data model.

Application: Endogenous network

- Without controlling for the endogeneity of the network

Parameters	CDSI Coef.	CDSI Marginal Effects		SART Coef.	SART Marginal Effects		SAR	
λ	0.443	0.363	(0.028)***	0.194	0.157	(0.005)***	0.185	(0.006)***
...								

- Controlling for the endogeneity of the network

Variance

Parameters	CDSI[1] Coef.	CDSI[1] Marginal Effects		SART Coef.	SART Marginal Effects		SAR	
λ	0.359	0.294	(0.028)***	0.173	0.141	(0.005)***	0.166	(0.006)***
$\rho\sigma_\varepsilon$	0.246	0.202	(0.011)***	0.253	0.205	(0.010)***	0.240	(0.013)***
$\bar{\rho}\sigma_\varepsilon$	0.202	0.166	(0.019)***	0.240	0.195	(0.018)***	0.218	(0.020)***

- Model with endogeneity is the best model according the likelihood ratio test.

Conclusion

- First model of random utility dealing with networks and count outcome.
- The model performs well on count data.
- Two main results.
  1. Integer nature of the outcome is important.
  2. The endogeneity of the network is important.
- (Next steps) Zeros inflated specification may be required (e.g., smoking).
- `CDatanet` package, https://github.com/ahoundetoungan/CDatanet.

```
CD <- CDnetNPL(formula = y ~ x1 + x2, contextual = TRUE,
               Glist = Network, optimizer = "nlm")
summary(CD)
```

Conclusion

- First model of random utility dealing with networks and count outcome.
- The model performs well on count data.
- Two main results.
  1. Integer nature of the outcome is important.
  2. The endogeneity of the network is important.
- (Next steps) Zeros inflated specification may be required (e.g., smoking).
- `CDatanet` package, https://github.com/ahoundetoungan/CDatanet.

```
CD <- CDnetNPL(formula = y ~ x1 + x2, contextual = TRUE,
               Glist = Network, optimizer = "nlm")
summary(CD)
```

Conclusion

- First model of random utility dealing with networks and count outcome.
- The model performs well on count data.
- Two main results.
  1. Integer nature of the outcome is important.
  2. The endogeneity of the network is important.
- (Next steps) Zeros inflated specification may be required (e.g., smoking).
- `CDatanet` package, https://github.com/ahoundetoungan/CDatanet.

```
CD <- CDnetNPL(formula = y ~ x1 + x2, contextual = TRUE,
               Glist = Network, optimizer = "nlm")
summary(CD)
```

Conclusion

- First model of random utility dealing with networks and count outcome.
- The model performs well on count data.
- Two main results.
  1. Integer nature of the outcome is important.
  2. The endogeneity of the network is important.
- (Next steps) Zeros inflated specification may be required (e.g., smoking).
- `CDatanet` package, https://github.com/ahoundetoungan/CDatanet.

```
CD <- CDnetNPL(formula = y ~ x1 + x2, contextual = TRUE,
               Glist = Network, optimizer = "nlm")
summary(CD)
```

Conclusion

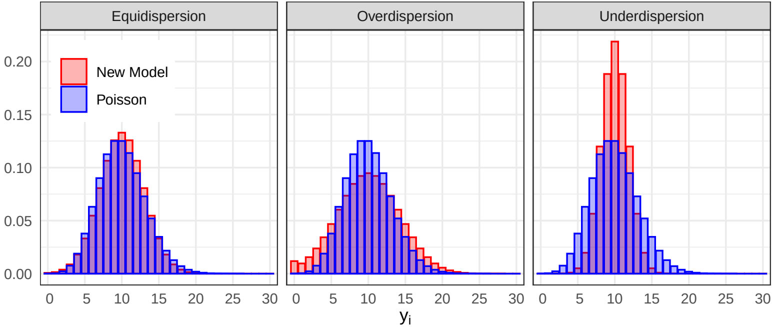
- First model of random utility dealing with networks and count outcome.
- The model performs well on count data.
- Two main results.
  1. Integer nature of the outcome is important.
  2. The endogeneity of the network is important.
- (Next steps) Zeros inflated specification may be required (e.g., smoking).
- `CDatanet` package, https://github.com/ahoundetoungan/CDatanet.

```
CD <- CDnetNPL(formula = y ~ x1 + x2, contextual = TRUE,
               Glist = Network, optimizer = "nlm")
summary(CD)
```

THANK YOU

Game: First Order Conditions (focs)

- Belief comparison with the standard Poisson model ($\lambda = 0$)

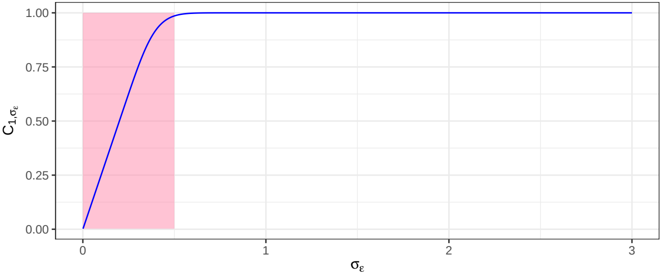


- Flexible model in term of dispersion fitting as the Generalized Poisson model.
  - The Poisson model only allows equidispersion;
  - The Negative Binomial model only allows overdispersion and equidispersion.

back

Game: Equilibrium

- Assume $\gamma = 1$ and $\mathbf{G}$ is row-normalized; ie $||\mathbf{G}||_\infty = 1$.
- Is the condition on λ much stronger than $|\lambda| < 1$?
- C_{1,σ_ε} (upper bound of λ when $\gamma = 1$ and $||\mathbf{G}||_\infty = 1$) as a function of σ_ε



- The condition $\sigma_\varepsilon < 0.5$ is likely to be violated in practice:
  - max of $\mathbf{Var}(y_i|\psi_i) < 0.34$;
  - Only two count choices concentrate more than 84% of observed data.

back

Variance of the two-stage estimation

- Unconditional variance

$$\mathbf{Var}(\hat{\theta}) = \mathbf{E}_u\left(\mathbf{Var}(\hat{\theta}|\tilde{\boldsymbol{\mu}})\right) + \mathbf{Var}_u\left(\mathbf{E}(\hat{\theta}|\tilde{\boldsymbol{\mu}})\right). \tag{12}$$

- Assumption: Let $\tilde{\boldsymbol{\mu}}_s$ be a draw of $\tilde{\boldsymbol{\mu}}$ from its posterior distribution and $\hat{\theta}_s$ be the estimator of θ_0 associated with $\tilde{\boldsymbol{\mu}}_s$. $\hat{\theta}_s$ is a consistent estimator of $\mathbf{E}(\hat{\theta}_s|\tilde{\boldsymbol{\mu}}_s)$.

$$\widehat{AsyVar}\left(\hat{\theta}_s\right) = \frac{1}{S}\sum_{s=1}^{S}\mathbf{Var}(\hat{\theta}_s|\tilde{\boldsymbol{\mu}}_s) + \frac{1}{S-1}\sum_{s=1}^{T}\left(\hat{\theta}_s - \hat{\hat{\theta}}\right)\left(\hat{\theta}_s - \hat{\hat{\theta}}\right)', \tag{13}$$

where $\tilde{\boldsymbol{\mu}}_1, \dots, \tilde{\boldsymbol{\mu}}_S$ are S draws of $\tilde{\boldsymbol{\mu}}$ with replacement from the population of the 10,000 simulations kept at the first stage, and $\hat{\hat{\theta}} = \frac{1}{S}\sum_{s=1}^{S}\hat{\theta}_s$. In practice, I set $S = 5,000$.

back