# Correcting for Sample Selection Bias in Dyadic Regressions

## An Application to Gravity Models


Gabriela Mara Miyazato Szini

*g.m.m.szini@uva.nl*

Supervised by: Prof. Dr. Frank Kleibergen


August 26, 2020

# Motivation

- Tinbergen [1962] established gravity equations: used to estimate models of international trade flows. Particularly, to infer the effects of trade barriers on flows.

# Motivation

- Tinbergen [1962] established gravity equations: used to estimate models of international trade flows. Particularly, to infer the effects of trade barriers on flows.

- **Before Helpman et al. [2008]:** studies only considered observations with positive trade flows when estimating the model, without imposing any special treatment on it.

- **Helpman et al. [2008]:** correct for sample selection using the standard Heckman [1979] approach. Gravity model

# Motivation

- Tinbergen [1962] established gravity equations: used to estimate models of international trade flows. Particularly, to infer the effects of trade barriers on flows.
- **Before Helpman et al. [2008]:** studies only considered observations with positive trade flows when estimating the model, without imposing any special treatment on it.
- **Helpman et al. [2008]:** correct for sample selection using the standard Heckman [1979] approach. Gravity model
- **What is the remaining problem?**
  The inclusion of two-way fixed effects when correcting for sample selection through Heckman [1979] approach leads to incidental parameter problem (Neyman and Scott [1948])

# Motivation

- Tinbergen [1962] established gravity equations: used to estimate models of international trade flows. Particularly, to infer the effects of trade barriers on flows.
- **Before Helpman et al. [2008]:** studies only considered observations with positive trade flows when estimating the model, without imposing any special treatment on it.
- **Helpman et al. [2008]:** correct for sample selection using the standard Heckman [1979] approach. Gravity model
- **What is the remaining problem?**
  The inclusion of two-way fixed effects when correcting for sample selection through Heckman [1979] approach leads to incidental parameter problem (Neyman and Scott [1948])
- **Main aim:** propose estimators that are not contaminated by this problem.

# Aim of the study

- Define a model to any dyadic data with sample selection.

# Aim of the study

- Define a model to any dyadic data with sample selection.
- Outline the incidental parameter problem that arises when employing the standard Heckman [1979] aproach.

# Aim of the study

- Define a model to any dyadic data with sample selection.
- Outline the incidental parameter problem that arises when employing the standard Heckman [1979] aproach.
- First stage of sample selection correction: method based on conditional likelihood estimators of Charbonneau [2017].

# Aim of the study

- Define a model to any dyadic data with sample selection.
- Outline the incidental parameter problem that arises when employing the standard Heckman [1979] aproach.
- First stage of sample selection correction: method based on conditional likelihood estimators of Charbonneau [2017].
- Propose new approaches to the second stage of sample selection correction:
  - Hybrid estimates of fixed effects + Lee's transformation.
  - Modification to Kyriazidou [1997].

# Aim of the study

- Define a model to any dyadic data with sample selection.
- Outline the incidental parameter problem that arises when employing the standard Heckman [1979] aproach.
- First stage of sample selection correction: method based on conditional likelihood estimators of Charbonneau [2017].
- Propose new approaches to the second stage of sample selection correction:
  - Hybrid estimates of fixed effects + Lee's transformation.
  - Modification to Kyriazidou [1997].
- Simulation exercise indicates that new approaches reduces the biases in the estimates.

# Model of dyadic interactions

$$y_{1,ij,t} = x'_{1,ij,t}\beta_1 + \vartheta_i + \chi_j + u_{ij,t} \quad \text{(Observation equation)}$$

$$y_{2,ij,t} = \mathbb{1}(y^{**}_{2,ij,t} > 0)$$

$$y^{**}_{2,ij,t} = x'_{2,ij,t}\beta^*_2 + \xi^*_i + \zeta^*_j + \eta^*_{ij,t} \quad \text{(Selection equation)}$$

$$(i = 1, ...N; j = 1, ...N, t = 1, ...T, i \neq j)$$

- $y_{1,ij,t}$ is observed only when $y_{2,ij,t} = 1$.
- $x_{1,ij,t}$ and $x_{2,ij,t}$ are vectors of $k$ and $q$ explanatory variables.
- $\beta_1 \in \mathbb{R}^k$ and $\beta^*_2 \in \mathbb{R}^q$
- $\chi_j$, $\vartheta_i$, $\xi^*_i$ and $\zeta^*_j$ are fixed effects.

# Model of dyadic interactions

$$y_{1,ij,t} = x'_{1,ij,t}\beta_1 + \vartheta_i + \chi_j + u_{ij,t} \quad \text{(Observation equation)}$$

$$y_{2,ij,t} = \mathbb{1}(y^{**}_{2,ij,t} > 0)$$

$$y^{**}_{2,ij,t} = x'_{2,ij,t}\beta^*_2 + \xi^*_i + \zeta^*_j + \eta^*_{ij,t} \quad \text{(Selection equation)}$$

$$(i = 1, ...N; j = 1, ...N, t = 1, ...T, i \neq j)$$

- $y_{1,ij,t}$ is observed only when $y_{2,ij,t} = 1$.
- $x_{1,ij,t}$ and $x_{2,ij,t}$ are vectors of $k$ and $q$ explanatory variables.
- $\beta_1 \in \mathbb{R}^k$ and $\beta^*_2 \in \mathbb{R}^q$
- $\chi_j$, $\vartheta_i$, $\xi^*_i$ and $\zeta^*_j$ are fixed effects.

**Main features:**

- Unobserved individual effects may depend on other variables arbitrarily.
- Errors of equations might be correlated.
- Directed links/outcomes.
- Selection equation indicates a network formation model with degree heterogeneity.

# Currently in the literature: Heckman's two-step approach

**Assumption:**

$$\mathbb{E}[u_{ij,t}\eta^*_{ij,t}] = \begin{cases} \sigma_{u\eta^*} \text{ if } i = i', j = j', t = t' \\ 0 \text{ otherwise} \end{cases}$$

We can rewrite the observation equation as:

$$\begin{aligned} &\mathbb{E}[y_{1,ij,t}|x_{1,ij,t}, \vartheta_i, \chi_j, y_{2,ij,t} = 1] \\ &= x'_{1,ij,t}\beta_1 + \vartheta_i + \chi_j + \mathbb{E}[u_{ij,t}|\eta^*_{ij,t} \geq \underbrace{-x'_{2,ij,t}\beta^*_2 - \xi^*_i - \zeta^*_j}_{z_{ij,t}}] \end{aligned}$$

# Currently in the literature: Heckman's two-step approach

**Assumption:**

$$\mathbb{E}[u_{ij,t}\eta^*_{ij,t}] = \begin{cases} \sigma_{u\eta^*} \text{ if } i = i', j = j', t = t' \\ 0 \text{ otherwise} \end{cases}$$

We can rewrite the observation equation as:

$$\begin{aligned} &\mathbb{E}[y_{1,ij,t}|x_{1,ij,t}, \vartheta_i, \chi_j, y_{2,ij,t} = 1] \\ &= x'_{1,ij,t}\beta_1 + \vartheta_i + \chi_j + \mathbb{E}[u_{ij,t}|\eta^*_{ij,t} \geq \underbrace{-x'_{2,ij,t}\beta^*_2 - \xi^*_i - \zeta^*_j}_{z_{ij,t}}] \end{aligned}$$

Further assuming that the conditional expectation in the last term is an unknown, smooth function $\varphi_{ij,t}$:

$$y_{1,ij,t} = x'_{1,ij,t}\beta_1 + \vartheta_i + \chi_j + \varphi_{ij,t}(z_{ij,t}) + \nu_{ij,t},$$

by construction, $\mathbb{E}[\nu_{ij,t}|x_{1,ij,t}, \vartheta_i, \chi_j, \varphi_{ij,t}(z_{ij,t}), y^{**}_{2,ij,t} > 0] = 0$.

# Currently in the literature: Heckman's two-step approach

**Assumption:**

$$\mathbb{E}[u_{ij,t}\eta^*_{ij,t}] = \begin{cases} \sigma_{u\eta^*} \text{ if } i = i', j = j', t = t' \\ 0 \text{ otherwise} \end{cases}$$

We can rewrite the observation equation as:

$$\begin{aligned} &\mathbb{E}[y_{1,ij,t}|x_{1,ij,t}, \vartheta_i, \chi_j, y_{2,ij,t} = 1] \\ &= x'_{1,ij,t}\beta_1 + \vartheta_i + \chi_j + \mathbb{E}[u_{ij,t}|\eta^*_{ij,t} \geq \underbrace{-x'_{2,ij,t}\beta^*_2 - \xi^*_i - \zeta^*_j}_{z_{ij,t}}] \end{aligned}$$

Further assuming that the conditional expectation in the last term is an unknown, smooth function $\varphi_{ij,t}$:

$$y_{1,ij,t} = x'_{1,ij,t}\beta_1 + \vartheta_i + \chi_j + \varphi_{ij,t}(z_{ij,t}) + \nu_{ij,t},$$

by construction, $\mathbb{E}[\nu_{ij,t}|x_{1,ij,t}, \vartheta_i, \chi_j, \varphi_{ij,t}(z_{ij,t}), y^{**}_{2,ij,t} > 0] = 0$.

**Identification:** either the function $\varphi_{ij,t}$ is specified or additional variables that satisfy exclusion restriction from the observation equation are included in $x_{2,ij,t}$.

# Currently in the literature: Heckman's two-step approach

Introducing functional form for $\varphi_{ij,t}$ through additional distributional assumptions:

**Assumption:**

The error term $u_{ij,t}$ is i.i.d. over $ij$ and $t$, such that $u_{ij,t} \sim N(0, \sigma_u^2)$.
The error term $\eta^*_{ij,t}$ is i.i.d. over $ij$ and $t$, such that $\eta^*_{ij,t} \sim N(0, 1)$.

# Currently in the literature: Heckman's two-step approach

Introducing functional form for $\varphi_{ij,t}$ through additional distributional assumptions:

**Assumption:**

The error term $u_{ij,t}$ is i.i.d. over $ij$ and $t$, such that $u_{ij,t} \sim N(0, \sigma_u^2)$.
The error term $\eta^*_{ij,t}$ is i.i.d. over $ij$ and $t$, such that $\eta^*_{ij,t} \sim N(0, 1)$.

From the standard results of the bivariate normal distribution:

$$y_{1,ij,t} = x'_{1,ij,t}\beta_1 + \vartheta_i + \chi_j + \sigma_{u\eta^*}\lambda_{ij,t}(z_{ij,t}) + \nu_{ij,t}$$

where: $\lambda_{ij,t}(z_{ij,t}) = \frac{\phi(z_{ij,t})}{1-\Phi(z_{ij,t})} = \frac{\phi(z_{ij,t})}{\Phi(-z_{ij,t})}$ is the inverse Mills-ratio.

# Currently in the literature: Heckman's two-step approach

Heckman [1979]'s approach in a nutshell:

- **Stage 1:** Estimate selection equation by MLE (probit).
- **Stage 2:** Obtain inverse Mills-ratio and estimate observation equation by FGLS.

# Currently in the literature: Heckman's two-step approach

Heckman [1979]'s approach in a nutshell:

- **Stage 1:** Estimate selection equation by MLE (probit).
- **Stage 2:** Obtain inverse Mills-ratio and estimate observation equation by FGLS.

Taking into account an **estimated value** of $\lambda_{ij,t}(z_{ij,t})$:

$$y_{1,ij,t} = x'_{1,ij,t}\beta_1 + \vartheta_i + \chi_j + \sigma_{u\eta^*}\hat{\lambda}_{ij,t} + \sigma_{u\eta^*}(\lambda_{ij,t} - \hat{\lambda}_{ij,t}) + \nu_{ij,t}$$

We can interpretate $\sigma_{u\eta^*}(\lambda_{ij,t} - \hat{\lambda}_{ij,t}) + \nu_{ij,t}$ as the error term of this equation.

# Currently in the literature: Heckman's two-step approach

Heckman [1979]'s approach in a nutshell:

- **Stage 1:** Estimate selection equation by MLE (probit).
- **Stage 2:** Obtain inverse Mills-ratio and estimate observation equation by FGLS.

Taking into account an **estimated value** of $\lambda_{ij,t}(z_{ij,t})$:

$$y_{1,ij,t} = x'_{1,ij,t}\beta_1 + \vartheta_i + \chi_j + \sigma_{u\eta^*}\hat{\lambda}_{ij,t} + \sigma_{u\eta^*}(\lambda_{ij,t} - \hat{\lambda}_{ij,t}) + \nu_{ij,t}$$

We can interpretate $\sigma_{u\eta^*}(\lambda_{ij,t} - \hat{\lambda}_{ij,t}) + \nu_{ij,t}$ as the error term of this equation.

**Theorem 1:**

The estimator of $\beta_1$ will be an asymptotically unbiased estimator, i.e., will converge to a limiting distribution centered around its true value only if $\mathbb{E}[(\lambda_{ij,t} - \hat{\lambda}_{ij,t})] = 0$.

Heckman's steps

# Which is the problem? The incidental parameter problem

**Fixed effects estimators in nonlinear models suffer from the incidental parameter problem (Neyman and Scott [1948])**.

# Which is the problem? The incidental parameter problem

**Fixed effects estimators in nonlinear models suffer from the incidental parameter problem (Neyman and Scott [1948])**.

**How does this problem arise?**
Denote:

$$\bar{\beta}_2^* = \underset{\beta_2^* \in \mathbb{R}^{\dim \beta_2^*}}{\arg\max}\ \mathbb{E}_\omega \Big[ \mathcal{L}(\beta_2^*, \hat{\omega}_{NN}^*(\beta_2^*)) \Big]$$

Using an asymptotic expansion for smooth likelihoods under appropriate regularity conditions (Fernández-Val and Weidner [2016]) Asymptotic expansion :

$$\bar{\beta}_2^* = \beta_{2,0}^* + \frac{\bar{B}_\infty}{(N-1)T} + \frac{\bar{D}_\infty}{(N-1)T} + o_P(((N-1)T)^{-1})$$

For some constants $\bar{B}_\infty$ and $\bar{D}_\infty$.

# Which is the problem? The incidental parameter problem

**Fixed effects estimators in nonlinear models suffer from the incidental parameter problem (Neyman and Scott [1948])**.

**How does this problem arise?**
Denote:

$$\bar{\beta}_2^* = \underset{\beta_2^* \in \mathbb{R}^{\dim \beta_2^*}}{\arg\max}\ \mathbb{E}_\omega\Big[\mathcal{L}(\beta_2^*, \hat{\omega}_{NN}^*(\beta_2^*))\Big]$$

Using an asymptotic expansion for smooth likelihoods under appropriate regularity conditions (Fernández-Val and Weidner [2016]) Asymptotic expansion :

$$\bar{\beta}_2^* = \beta_{2,0}^* + \frac{\bar{B}_\infty}{(N-1)T} + \frac{\bar{D}_\infty}{(N-1)T} + o_P(((N-1)T)^{-1})$$

For some constants $\bar{B}_\infty$ and $\bar{D}_\infty$. By the properties of the maximum likelihood estimator:

$$\sqrt{N(N-1)T}(\hat{\beta}_2^* - \bar{\beta}_2^*) \xrightarrow{d} N(0, \bar{V}_{B\infty})$$

For some $\bar{V}_{B\infty}$. By Slutsky Theorem:

$$\sqrt{N(N-1)T}(\hat{\beta}_2^* - \beta_{2,0}^*) \xrightarrow{d} N\Big(\frac{\bar{B}_\infty}{\sqrt{T}} + \frac{\bar{D}_\infty}{\sqrt{T}}, \bar{V}_{B\infty}\Big)$$

# Which is the problem? The incidental parameter problem

As $N \to \infty$ and considering fixed $T$:

- $\hat{\beta}_2^* \xrightarrow{p} \beta_{2,0}^*$ : estimates are consistent.
- **But** the estimates converge to a distribution that is not centered at the true parameter value, which leads to incorrect asymptotic confidence intervals.

# Which is the problem? The incidental parameter problem

As $N \to \infty$ and considering fixed $T$:

- $\hat{\beta}_2^* \xrightarrow{p} \beta_{2,0}^*$ : estimates are consistent.
- **But** the estimates converge to a distribution that is not centered at the true parameter value, which leads to incorrect asymptotic confidence intervals.

The source of the problem is that the estimates of the nuisance parameters converge at a slower rate than the structural parameters.

# Which is the problem? The incidental parameter problem

As $N \to \infty$ and considering fixed $T$:

- $\hat{\beta}_2^* \xrightarrow{p} \beta_{2,0}^*$ : estimates are consistent.
- **But** the estimates converge to a distribution that is not centered at the true parameter value, which leads to incorrect asymptotic confidence intervals.

The source of the problem is that the estimates of the nuisance parameters converge at a slower rate than the structural parameters.

**Conjecture 1:**

The asymptotic bias in the estimates of $\beta_2^*$ carries over both to the estimates of the fixed effects, $\hat{\omega}_{NN}^*$ and to the estimates of the inverse Mills-ratio, $\hat{\lambda}_{ij,t}$.

**Conditions of Theorem 1 do not hold.**

# Our proposed methods: Outline

1. **First stage: conditional logit estimation of Charbonneau [2017].**
   - **Differences out the fixed effects $\rightarrow$ free of incidental parameter problem.**
2. Second stage: given estimates of first stage, we propose two methods.
   - Hybrid approach to retrieve FE of first stage + Lee's transformation to apply Heckman even when errors are logistic distributed.
   - A modification to the Kyriazidou [1997] approach that relies on the idea of differencing out the selection effects.

# Conditional logit estimation of Charbonneau [2017]

**Main idea:** difference out two-way fixed effects in selection equation.
**Result:** obtain estimates of $\beta_2^*$ that are not contaminated by the incidental parameters problem.

# Conditional logit estimation of Charbonneau [2017]

**Main idea:** difference out two-way fixed effects in selection equation.
**Result:** obtain estimates of $\beta_2^*$ that are not contaminated by the incidental parameters problem.

Modify Assumption on the distribution of $\eta_{ij}^*$ to:

**Assumption:**

The idiosyncratic terms $\eta_{ij}^*$ are i.i.d. across $ij$ and follow a standard logistic distribution conditional on the regressors and fixed effects.

For simplicity: consider only one period $t$.

# Conditional logit estimation of Charbonneau [2017]

**Main idea:** difference out two-way fixed effects in selection equation.
**Result:** obtain estimates of $\beta_2^*$ that are not contaminated by the incidental parameters problem.

Modify Assumption on the distribution of $\eta_{ij}^*$ to:

**Assumption:**

The idiosyncratic terms $\eta_{ij}^*$ are i.i.d. across $ij$ and follow a standard logistic distribution conditional on the regressors and fixed effects.

For simplicity: consider only one period $t$.

Assumption implies that:

$$\Pr(y_{2,ij} = 1 | x_2, \xi^*, \zeta^*, \beta_2^*) = \frac{\exp(x_{2,ij}' \beta_2^* + \xi_i^* + \zeta_j^*)}{1 + \exp(x_{2,ij}' \beta_2^* + \xi_i^* + \zeta_j^*)}$$

where $x_2$ is the vector of observations $x_{2,ij}$ for all possible pairs $ij$.

# Conditional logit estimation of Charbonneau [2017]

Fix distinct indices $i$, $j$, $k$ and $l$ from the set formed by $i' = 1, ...N$. Then:

$$\Pr(y_{2,lj} = 1 | x_2, \xi^*, \zeta^*, \beta_2^*, y_{2,lk} + y_{2,lj} = 1, y_{2,ij} + y_{2,ik} = 1, y_{2,ij} + y_{2,lj} = 1)$$
$$= \frac{\exp(((x_{2,lj} - x_{2,lk}) - (x_{2,ij} - x_{2,ik}))'\beta_2^*)}{1 + \exp(((x_{2,lj} - x_{2,lk}) - (x_{2,ij} - x_{2,ik}))'\beta_2^*)}$$

# Conditional logit estimation of Charbonneau [2017]

Fix distinct indices $i$, $j$, $k$ and $l$ from the set formed by $i' = 1, ...N$. Then:

$$\Pr(y_{2,lj} = 1 | x_2, \xi^*, \zeta^*, \beta_2^*, y_{2,lk} + y_{2,lj} = 1, y_{2,ij} + y_{2,ik} = 1, y_{2,ij} + y_{2,lj} = 1)$$
$$= \frac{\exp(((x_{2,lj} - x_{2,lk}) - (x_{2,ij} - x_{2,ik}))'\beta_2^*)}{1 + \exp(((x_{2,lj} - x_{2,lk}) - (x_{2,ij} - x_{2,ik}))'\beta_2^*)}$$

**Which no longer depends on fixed effects.**

# Conditional logit estimation of Charbonneau [2017]

Fix distinct indices $i$, $j$, $k$ and $l$ from the set formed by $i' = 1, ...N$. Then:

$$\Pr(y_{2,lj} = 1 | x_2, \xi^*, \zeta^*, \beta_2^*, y_{2,lk} + y_{2,lj} = 1, y_{2,ij} + y_{2,ik} = 1, y_{2,ij} + y_{2,lj} = 1)$$
$$= \frac{\exp(((x_{2,lj} - x_{2,lk}) - (x_{2,ij} - x_{2,ik}))'\beta_2^*)}{1 + \exp(((x_{2,lj} - x_{2,lk}) - (x_{2,ij} - x_{2,ik}))'\beta_2^*)}$$

**Which no longer depends on fixed effects.**

Define the variables:

$$z = \frac{(y_{2,lj} - y_{2,lk}) - (y_{2,ij} - y_{2,ik})}{2}$$
$$r = (x_{2,lj} - x_{2,lk}) - (x_{2,ij} - x_{2,ik}))$$

Event that $z \in \{-1, 1\}$ corresponds to the condition that for any $ij$, $l$ and $k$ satisfies $y_{2,lk} + y_{2,lj} = 1, y_{2,ij} + y_{2,ik} = 1, y_{2,ij} + y_{2,lj} = 1$.

# Conditional logit estimation: Estimator

**Conditional on $x_2$ and $z \in \{-1, 1\}$, the distribution of $z$ is logistic and does not depend on the fixed effects.**

- When $z = 1$, we have necessarily that $y_{2,lj} = 1$, and when $z = -1$, it is necessarily zero.

# Conditional logit estimation: Estimator

**Conditional on $x_2$ and $z \in \{-1, 1\}$, the distribution of $z$ is logistic and does not depend on the fixed effects.**

- When $z = 1$, we have necessarily that $y_{2,lj} = 1$, and when $z = -1$, it is necessarily zero.

**Main features:**

- Objective function has the functional form of a standard logistic regression, but **asymptotic variance is of a sandwich form**.
- Controls for degree heterogeneity.
- Allows for **sparse networks** (fixed effects need not to be estimated).
- Possible to **test for homophily**.
- **Drawback :** not all quadruples are informative.

Asymptotic properties

# Our proposed methods: Outline

1. First stage: conditional logit estimation of Charbonneau [2017].
   - Differences out the fixed effects $\rightarrow$ free of incidental parameter problem.

2. **Second stage: given estimates of first stage, we propose two methods.**
   - **Hybrid approach to retrieve FE of first stage + Lee's transformation to apply Heckman even when errors are logistic distributed.**
   - A modification to the Kyriazidou [1997] approach that relies on the idea of differencing out the selection effects.

# First approach: Hybrid estimates

Wooldridge et al. [2018]: the fixed effects $\xi^*$ and $\zeta^*$ are estimated from the unconditional maximum likelihood when the slope parameters are set to $\hat{\beta}^*_{2,CL}$.

$$\hat{\omega}^*_{NN}(\hat{\beta}^*_{2,CL}) = \underset{\omega^*_{NN} \in \mathbb{R}^{\dim \omega^*_{NN}}}{\arg\max} \; \mathcal{L}(\hat{\beta}^*_{2,CL}, \omega^*_{NN})$$

where:

$$\begin{aligned} &\mathcal{L}(\hat{\beta}^*_{2,CL}, \omega^*_{NN}) \\ &= (N(N-1))^{-1} \Big\{ \sum_{i=1}^{N} \sum_{j \neq i} \log f_{Y_2}(y_{2,ij} | x_{2,ij}, \xi^*, \zeta^*, \beta^*_2) - b(\iota'_{NN} \omega^*_{NN})^2 / 2 \Big\} \end{aligned}$$

$$\begin{aligned} f_{Y_2}(y_{2,ij} | x_{2,ij}, \xi^*, \zeta^*, \beta^*_2) = &\; F(x'_{2,ij} \hat{\beta}^*_{2,CL} + \xi^*_i + \zeta^*_j)^{y_{2,ij}} \\ &\times [1 - F(x'_{2,ij} \hat{\beta}^*_{2,CL} + \xi^*_i + \zeta^*_j)]^{1 - y_{2,ij}}, \end{aligned}$$

with: $\omega^*_{NN}$ a vector that collects the fixed effects, $F$ a standard logistic distribution function, $b > 0$ is an arbitrary constant, $\iota_{NN} = (1'_N, -1'_N)'$ and $1_N$ denotes a vector of ones of dimension $N$.

# First approach: Hybrid estimates

Wooldridge et al. [2018]: the fixed effects $\xi^*$ and $\zeta^*$ are estimated from the unconditional maximum likelihood when the slope parameters are set to $\hat{\beta}^*_{2,CL}$.

$$\hat{\omega}^*_{NN}(\hat{\beta}^*_{2,CL}) = \underset{\omega^*_{NN}\in\mathbb{R}^{\dim \omega^*_{NN}}}{\arg\max}\ \mathcal{L}(\hat{\beta}^*_{2,CL}, \omega^*_{NN})$$

where:

$$\begin{aligned}&\mathcal{L}(\hat{\beta}^*_{2,CL}, \omega^*_{NN}) \\ &= (N(N-1))^{-1}\Big\{\sum_{i=1}^{N}\sum_{j\neq i} \log f_{Y_2}(y_{2,ij}|x_{2,ij}, \xi^*, \zeta^*, \beta^*_2) - b(\iota'_{NN}\omega^*_{NN})^2/2\Big\}\end{aligned}$$

$$\begin{aligned}f_{Y_2}(y_{2,ij}|x_{2,ij}, \xi^*, \zeta^*, \beta^*_2) &= F(x'_{2,ij}\hat{\beta}^*_{2,CL} + \xi^*_i + \zeta^*_j)^{y_{2,ij}} \\ &\times [1 - F(x'_{2,ij}\hat{\beta}^*_{2,CL} + \xi^*_i + \zeta^*_j)]^{1-y_{2,ij}},\end{aligned}$$

with: $\omega^*_{NN}$ a vector that collects the fixed effects, $F$ a standard logistic distribution function, $b > 0$ is an arbitrary constant, $\iota_{NN} = (1'_N, -1'_N)'$ and $1_N$ denotes a vector of ones of dimension $N$.

**We should also control for the perfect prediction problem.**

# First approach: Lee's transformation

First, we rewrite the observation equation as:

$$y_{1,ij,t} = x'_{1,ij}\beta_1 + \vartheta_i + \chi_j + \sigma_u u^*_{ij}$$

where $u^*_{ij} \sim N(0,1)$ and $\sigma_u > 0$.

# First approach: Lee's transformation

First, we rewrite the observation equation as:

$$y_{1,ij,t} = x'_{1,ij}\beta_1 + \vartheta_i + \chi_j + \sigma_u u^*_{ij}$$

where $u^*_{ij} \sim N(0,1)$ and $\sigma_u > 0$.

Conditional on $x_1, x_2, \vartheta_i, \chi_j, \xi^*_i, \zeta^*_j$ the errors have continuous distribution functions $\Phi(u^*)$ and $F(\eta^*)$, and are **allowed to be correlated, suposing a correlation of** $\rho$.

# First approach: Lee's transformation

First, we rewrite the observation equation as:

$$y_{1,ij,t} = x'_{1,ij}\beta_1 + \vartheta_i + \chi_j + \sigma_u u^*_{ij}$$

where $u^*_{ij} \sim N(0,1)$ and $\sigma_u > 0$.

Conditional on $x_1, x_2, \vartheta_i, \chi_j, \xi^*_i, \zeta^*_j$ the errors have continuous distribution functions $\Phi(u^*)$ and $F(\eta^*)$, and are **allowed to be correlated, suposing a correlation of** $\rho$.

**Main idea of Lee [1983]:** suggest a proper bivariate distribution that can be applied to the Heckman method with the specified marginal distributions.

# First approach: Lee's transformation

1. Apply a transformation to the error term such that it is distributed as a standard normal: $\eta^{**} = J(\eta^*) = \Phi^{-1}(F(\eta^*))$

# First approach: Lee's transformation

1. Apply a transformation to the error term such that it is distributed as a standard normal: $\eta^{**} = J(\eta^*) = \Phi^{-1}(F(\eta^*))$

**Assumption:**

The transformed random variables $\eta^{**}$ and $u^*$ are jointly normally distributed with zero means, unit variances and correlation of $\rho$.

# First approach: Lee's transformation

1. Apply a transformation to the error term such that it is distributed as a standard normal: $\eta^{**} = J(\eta^*) = \Phi^{-1}(F(\eta^*))$

**Assumption:**

The transformed random variables $\eta^{**}$ and $u^*$ are jointly normally distributed with zero means, unit variances and correlation of $\rho$.

2. A proper bivariate distribution of $(\eta^*, u^*)$, denoted by $H$, is derived such that:

$$H(\eta^*, u^*, \rho) = B(J(\eta^*), u^*; \rho)$$

# First approach: Lee's transformation

1. Apply a transformation to the error term such that it is distributed as a standard normal: $\eta^{**} = J(\eta^*) = \Phi^{-1}(F(\eta^*))$

**Assumption:**

The transformed random variables $\eta^{**}$ and $u^*$ are jointly normally distributed with zero means, unit variances and correlation of $\rho$.

2. A proper bivariate distribution of $(\eta^*, u^*)$, denoted by $H$, is derived such that:

$$H(\eta^*, u^*, \rho) = B(J(\eta^*), u^*; \rho)$$

3. $F(\cdot)$ is absolutely continuous, thus, the transformation $J(\cdot) = \Phi^{-1}(F(\cdot))$ is strictly increasing, such that $y_{2,ij} = 1$ if and only if:

$$J(-x'_{2,ij}\beta_2^* - \xi_i^* - \zeta_j^*) < J(\eta_{ij}^*)$$

# First approach: Lee's transformation

Then, the previous model is statistically equivalent to:

$$y_{1,ij,t} = x'_{1,ij}\beta_1 + \vartheta_i + \chi_j + \sigma_u u^*_{ij}$$
$$y^{***}_{2,ij} = J(x'_{2,ij}\beta^*_2 + \xi^*_i + \zeta^*_j) + \eta^{**}_{ij}$$

where $y^{***}_{2,ij} = J(y^{**}_{2,ij})$.

# First approach: Lee's transformation

Then, the previous model is statistically equivalent to:

$$y_{1,ij,t} = x'_{1,ij}\beta_1 + \vartheta_i + \chi_j + \sigma_u u^*_{ij}$$
$$y^{***}_{2,ij} = J(x'_{2,ij}\beta^*_2 + \xi^*_i + \zeta^*_j) + \eta^{**}_{ij}$$

where $y^{***}_{2,ij} = J(y^{**}_{2,ij})$. We can apply the Heckman approach to this model, with the inverse Mills-ratio:

$$\lambda_{ij}(z_{ij}) = \frac{\phi(J(z_{ij}))}{1 - \Phi(J(z_{ij}))} = \frac{\phi(J(z_{ij}))}{F(-z_{ij})}$$

# First approach: Lee's transformation

Then, the previous model is statistically equivalent to:

$$
\begin{aligned}
y_{1,ij,t} &= x'_{1,ij}\beta_1 + \vartheta_i + \chi_j + \sigma_u u^*_{ij} \\
y^{***}_{2,ij} &= J(x'_{2,ij}\beta^*_2 + \xi^*_i + \zeta^*_j) + \eta^{**}_{ij}
\end{aligned}
$$

where $y^{***}_{2,ij} = J(y^{**}_{2,ij})$. We can apply the Heckman approach to this model, with the inverse Mills-ratio:

$$
\lambda_{ij}(z_{ij}) = \frac{\phi(J(z_{ij}))}{1 - \Phi(J(z_{ij}))} = \frac{\phi(J(z_{ij}))}{F(-z_{ij})}
$$

Equation to be estimated by OLS (or FGLS):

$$
y_{1,ij} = x'_{1,ij}\beta_1 + \vartheta_i + \chi_j + (\sigma_u\rho)\lambda_{ij}(z_{ij}) + \nu_{ij}
$$

By construction : $\mathbb{E}[\nu_{ij}|x_{1,ij}, \vartheta_i, \chi_j, \lambda_{ij}(z_{ij}), y^{***}_{2,ij} > 0] = 0$.

# Our proposed methods: Outline

1. First stage: conditional logit estimation of Charbonneau [2017].
   - Differences out the fixed effects $\rightarrow$ free of incidental parameter problem.
2. **Second stage: given estimates of first stage, we propose two methods.**
   - Hybrid approach to retrieve FE of first stage + Lee's transformation to apply Heckman even when errors are logistic distributed.
   - **A modification to the Kyriazidou [1997] approach that relies on the idea of differencing out the selection effects.**

# Second approach: modification to Kyriazidou [1997]

Denote, for the quadruple $ij$, $ik$, $lj$, $lk$:

$$\varsigma_{ijlk} \equiv (x_{1,ij}, x_{2,ij}, x_{1,ik}, x_{2,ik}, x_{1,lj}, x_{2,lj}, x_{1,lk}, x_{2,lk}, \xi_i^*, \xi_l^*, \zeta_j^*, \zeta_k^*, \vartheta_i, \vartheta_l, \chi_j, \chi_k)$$

Taking the differences defined by:

$$\Delta_{ijlk} y_{1,ij} = (y_{1,ij} - y_{1,ik}) - (y_{1,lj} - y_{1,lk})$$

# Second approach: modification to Kyriazidou [1997]

Denote, for the quadruple $ij$, $ik$, $lj$, $lk$:

$$\varsigma_{ijlk} \equiv (x_{1,ij}, x_{2,ij}, x_{1,ik}, x_{2,ik}, x_{1,lj}, x_{2,lj}, x_{1,lk}, x_{2,lk}, \xi_i^*, \xi_l^*, \zeta_j^*, \zeta_k^*, \vartheta_i, \vartheta_l, \chi_j, \chi_k)$$

Taking the differences defined by:

$$\Delta_{ijlk} y_{1,ij} = (y_{1,ij} - y_{1,ik}) - (y_{1,lj} - y_{1,lk})$$

Fixed effects in observation can be easily differenced out:

$$\begin{aligned} &\mathbb{E}[\Delta_{ijlk} y_{1,ij} | y_{2,ij} = y_{2,ik} = y_{2,lj} = y_{2,lk} = 1, \varsigma_{ijlk}] \\ &= \Delta_{ijlk} x_{1,ij}' \beta_1 + \mathbb{E}(\Delta_{ijlk} u_{1,ij} | y_{2,ij} = y_{2,ik} = y_{2,lj} = y_{2,lk} = 1, \varsigma_{ijlk}) \end{aligned}$$

# Second approach: modification to Kyriazidou [1997]

Denote, for the quadruple $ij$, $ik$, $lj$, $lk$:

$$\varsigma_{ijlk} \equiv (x_{1,ij}, x_{2,ij}, x_{1,ik}, x_{2,ik}, x_{1,lj}, x_{2,lj}, x_{1,lk}, x_{2,lk}, \xi_i^*, \xi_l^*, \zeta_j^*, \zeta_k^*, \vartheta_i, \vartheta_l, \chi_j, \chi_k)$$

Taking the differences defined by:

$$\Delta_{ijlk} y_{1,ij} = (y_{1,ij} - y_{1,ik}) - (y_{1,lj} - y_{1,lk})$$

Fixed effects in observation can be easily differenced out:

$$\begin{aligned} &\mathbb{E}[\Delta_{ijlk} y_{1,ij} | y_{2,ij} = y_{2,ik} = y_{2,lj} = y_{2,lk} = 1, \varsigma_{ijlk}] \\ &= \Delta_{ijlk} x_{1,ij}' \beta_1 + \mathbb{E}(\Delta_{ijlk} u_{1,ij} | y_{2,ij} = y_{2,ik} = y_{2,lj} = y_{2,lk} = 1, \varsigma_{ijlk}) \end{aligned}$$

**But, sample selection effects remain.**
To be differenced out: $\mathbb{E}(\Delta_{ijlk} u_{1,ij} | y_{2,ij} = y_{2,ik} = y_{2,lj} = y_{2,lk} = 1, \varsigma_{ijlk}) = 0$

# Second approach: modification to Kyriazidou [1997]

- To be differenced out it boils down to:

$$\mathbb{E}((u_{ij} - u_{ik}) - (u_{lj} - u_{lk}))|y_{2,ij} = y_{2,ik} = y_{2,lj} = y_{2,lk} = 1, \varsigma_{ijlk}) = 0$$

# Second approach: modification to Kyriazidou [1997]

- To be differenced out it boils down to:

$$\mathbb{E}((u_{ij} - u_{ik}) - (u_{lj} - u_{lk}))|y_{2,ij} = y_{2,ik} = y_{2,lj} = y_{2,lk} = 1, \varsigma_{ijlk}) = 0$$

- We can denote: $\varphi_{ijlk} = \mathbb{E}(u_{ij}|y_{2,ij} = y_{2,ik} = y_{2,lj} = y_{2,lk} = 1, \varsigma_{ijlk})$

# Second approach: modification to Kyriazidou [1997]

- To be differenced out it boils down to:

$$\mathbb{E}((u_{ij} - u_{ik}) - (u_{lj} - u_{lk}))|y_{2,ij} = y_{2,ik} = y_{2,lj} = y_{2,lk} = 1, \varsigma_{ijlk}) = 0$$

- We can denote: $\varphi_{ijlk} = \mathbb{E}(u_{ij}|y_{2,ij} = y_{2,ik} = y_{2,lj} = y_{2,lk} = 1, \varsigma_{ijlk})$

- For dyad $ij$, due to errors being i.i.d., it is a function of the single index : $z_{ij} = -x'_{2,ij}\hat{\beta}^*_{CL,2} - \xi^*_i - \zeta^*_j$

# Second approach: modification to Kyriazidou [1997]

- To be differenced out it boils down to:

$$\mathbb{E}((u_{ij} - u_{ik}) - (u_{lj} - u_{lk}))|y_{2,ij} = y_{2,ik} = y_{2,lj} = y_{2,lk} = 1, \varsigma_{ijlk}) = 0$$

- We can denote: $\varphi_{ijlk} = \mathbb{E}(u_{ij}|y_{2,ij} = y_{2,ik} = y_{2,lj} = y_{2,lk} = 1, \varsigma_{ijlk})$

- For dyad $ij$, due to errors being i.i.d., it is a function of the single index : $z_{ij} = -x'_{2,ij}\hat{\beta}^*_{CL,2} - \xi^*_i - \zeta^*_j$

- If the difference between the single indices:

$$(z_{ij} - z_{ik}) - (z_{lj} - z_{lk})$$

  is close to zero, if $\varphi_{ijlk}$ is a smooth enough function, then $\Delta_{ijlk}\varphi_{ijlk}$ is also close to zero.

# Second approach: modification to Kyriazidou [1997]

- To be differenced out it boils down to:

$$\mathbb{E}((u_{ij} - u_{ik}) - (u_{lj} - u_{lk}))|y_{2,ij} = y_{2,ik} = y_{2,lj} = y_{2,lk} = 1, \varsigma_{ijlk}) = 0$$

- We can denote: $\varphi_{ijlk} = \mathbb{E}(u_{ij}|y_{2,ij} = y_{2,ik} = y_{2,lj} = y_{2,lk} = 1, \varsigma_{ijlk})$

- For dyad $ij$, due to errors being i.i.d., it is a function of the single index : $z_{ij} = -x'_{2,ij}\hat{\beta}^*_{CL,2} - \xi^*_i - \zeta^*_j$

- If the difference between the single indices:

$$(z_{ij} - z_{ik}) - (z_{lj} - z_{lk})$$

  is close to zero, if $\varphi_{ijlk}$ is a smooth enough function, then $\Delta_{ijlk}\varphi_{ijlk}$ is also close to zero.

- The magnitude of the difference between single indices is free of fixed effects, and boils down to $\Delta_{ijlk}x'_{2,ij}\hat{\beta}^*_{2,CL}$.

# Second approach: modification to Kyriazidou [1997]

- To be differenced out it boils down to:

$$\mathbb{E}((u_{ij} - u_{ik}) - (u_{lj} - u_{lk}))|y_{2,ij} = y_{2,ik} = y_{2,lj} = y_{2,lk} = 1, \varsigma_{ijlk}) = 0$$

- We can denote: $\varphi_{ijlk} = \mathbb{E}(u_{ij}|y_{2,ij} = y_{2,ik} = y_{2,lj} = y_{2,lk} = 1, \varsigma_{ijlk})$

- For dyad $ij$, due to errors being i.i.d., it is a function of the single index : $z_{ij} = -x'_{2,ij}\hat{\beta}^*_{CL,2} - \xi^*_i - \zeta^*_j$

- If the difference between the single indices:

$$(z_{ij} - z_{ik}) - (z_{lj} - z_{lk})$$

  is close to zero, if $\varphi_{ijlk}$ is a smooth enough function, then $\Delta_{ijlk}\varphi_{ijlk}$ is also close to zero.

- The magnitude of the difference between single indices is free of fixed effects, and boils down to $\Delta_{ijlk}x'_{2,ij}\hat{\beta}^*_{2,CL}$.

**Main idea:** assign weights for observations according to how close to zero $|\Delta_{ijlk}x'_{2,ij}\hat{\beta}^*_{2,CL}|$ is.

# Second approach: modification to Kyriazidou [1997]

The estimator becomes:

$$\hat{\beta}_1 = \Big[ \sum_{i=1}^{N} \sum_{j \neq i} \sum_{k \neq i,j} \sum_{l \neq i,j,k} \hat{\Psi}_{ijkl} \Delta_{ijkl} x'_{1,ij} \Delta_{ijkl} x_{1,ij} \Phi_{ijkl} \Big]^{-1}$$

$$\Big[ \sum_{i=1}^{N} \sum_{j \neq i} \sum_{k \neq i,j} \sum_{l \neq i,j,k} \hat{\Psi}_{ijkl} \Delta_{ijkl} x'_{1,ij} \Delta_{ijkl} y_{1,ijkl} \Phi_{ijkl} \Big]$$

with $\Phi_{ijkl} = \mathbb{1}(y_{2,ij} = y_{2,ik} = y_{2,lj} = y_{2,lk} = 1)$, and weights:

$$\hat{\Psi}_{ijkl} = \frac{1}{h_n} K \Big( \frac{\Delta_{ijkl} x'_{2,ij} \hat{\beta}^{*}_{2,CL}}{h_n} \Big)$$

where $K$ is a kernel density function, and $h_n$ is a sequence of bandwidths which tends to zero as $N(N-1) \to \infty$.

# Second approach: modification to Kyriazidou [1997]

The estimator becomes:

$$\hat{\beta}_1 = \Big[ \sum_{i=1}^{N} \sum_{j \neq i} \sum_{k \neq i,j} \sum_{l \neq i,j,k} \hat{\Psi}_{ijkl} \Delta_{ijkl} x'_{1,ij} \Delta_{ijkl} x_{1,ij} \Phi_{ijkl} \Big]^{-1}$$

$$\Big[ \sum_{i=1}^{N} \sum_{j \neq i} \sum_{k \neq i,j} \sum_{l \neq i,j,k} \hat{\Psi}_{ijkl} \Delta_{ijkl} x'_{1,ij} \Delta_{ijkl} y_{1,ijkl} \Phi_{ijkl} \Big]$$

with $\Phi_{ijkl} = \mathbb{1}(y_{2,ij} = y_{2,ik} = y_{2,lj} = y_{2,lk} = 1)$, and weights:

$$\hat{\Psi}_{ijkl} = \frac{1}{h_n} K\Big( \frac{\Delta_{ijkl} x'_{2,ij} \hat{\beta}^*_{2,CL}}{h_n} \Big)$$

where $K$ is a kernel density function, and $h_n$ is a sequence of bandwidths which tends to zero as $N(N-1) \to \infty$.

For a given order of differentiability $r$ of the expression $\mathbb{E}(\Delta x'_1 \varphi \Phi | \Delta x'_2 \beta^*_2)$, $h_n = h(N(N-1))^{-\mu}$ should be chosen such that $\mu = 1/(2(r+1)+1)$.
**Problem of choosing a bandwith boils down to choosing a constant $h$.**

Optimal $h$

# Simulations: Aim

Compare in the first stage (selection equation) the approaches:

- Probit with dummies for fixed effects
- Probit with dummies for fixed effects correcting for perfect prediction
- Charbonneau

Compare in the second stage (observation equation) the approaches:

- Standard Heckman
- Standard Heckman correcting for perfect prediction
- Hybrid + Lee's transformation
- Modified Kyriazidou

Assume that there is no misspecification in the distributional assumptions and in the parametric form of models.

500 Monte Carlo simulations, sample size of $N = 25$, yielding $(N(N-1)) = 600$ observations.

# Simulations: DGP Designs

Considering only one time period:

$$
\begin{aligned}
y_{1,ij} &= x_{1,ij}\beta_{11} + x_{2,ij}\beta_{12} + \vartheta_i + \chi_j + u_{ij} \\
y_{2,ij} &= \mathbb{1}(y^{**}_{2,ij} > 0) \\
y^{**}_{2,ij} &= x_{1,ij}\beta^*_{21} + x_{2,ij}\beta^*_{22} + x_{3,ij}\beta^*_{23} + \xi^*_i + \zeta^*_j + \eta^*_{ij}
\end{aligned}
$$

$$(i = 1, ...N; j = 1, ...N, i \neq j)$$

Set: $\beta_{11} = 1$, $\beta_{12} = 2.5$, $\beta^*_{21} = 0.8$, $\beta^*_{22} = 1$ and $\beta^*_{23} = 2$.

# Simulations: DGP Designs

Considering only one time period:

$$y_{1,ij} = x_{1,ij}\beta_{11} + x_{2,ij}\beta_{12} + \vartheta_i + \chi_j + u_{ij}$$
$$y_{2,ij} = \mathbb{1}(y^{**}_{2,ij} > 0)$$
$$y^{**}_{2,ij} = x_{1,ij}\beta^*_{21} + x_{2,ij}\beta^*_{22} + x_{3,ij}\beta^*_{23} + \xi^*_i + \zeta^*_j + \eta^*_{ij}$$
$$(i = 1, ...N; j = 1, ...N, i \neq j)$$

Set: $\beta_{11} = 1$, $\beta_{12} = 2.5$, $\beta^*_{21} = 0.8$, $\beta^*_{22} = 1$ and $\beta^*_{23} = 2$.

**Design 5:** $\vartheta_i = \chi_j = 0$, $\xi^*_i$ and $\zeta^*_j$ drawn from standard normal distribution.

- $x_1 = rndn + \xi^*_i + \zeta^*_j$, where *rndn* is a standard normal.
- $x_2 = \mathbb{1}\{rndn \leq 0.5\}$, being a bivariate variable.
- $x_3 = \mathbb{1}\{rndn \leq 0.5\}$, being a bivariate variable.

**Design 6:** $\xi^*_i = \zeta^*_j = 0$, $\vartheta_i$ and $\chi_j$ drawn from standard normal distribution.
$x_1 = rndn + \vartheta_i + \chi_j$.

**Design 7:** All fixed effects and $x_1 = rndn + \vartheta_i + \chi_j + \xi^*_i + \zeta^*_j$.

# Simulations: First stage *t-tests* for structural parameters

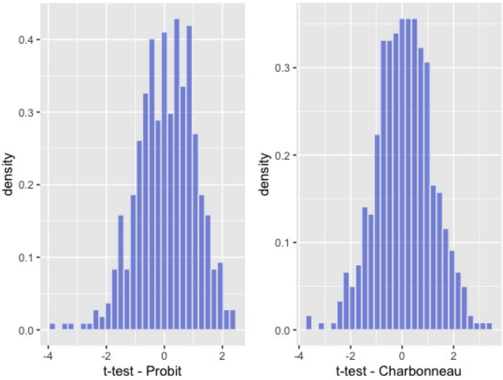


Figure 1: Histogram of the *t-test* for estimated $\beta^{*}_{21}$ in Design 6

Design that do not contain fixed effects in selection equation both approaches deliver similar results: distributions are centered around zero.

# Simulations: First stage *t-tests* for structural parameters

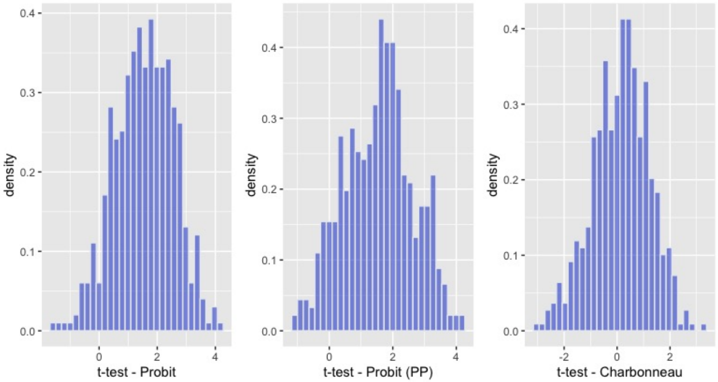


Figure 2: Histogram of the *t-test* for estimated $\beta^{*}_{21}$ in Design 5

# Simulations: First stage *t-tests* for structural parameters

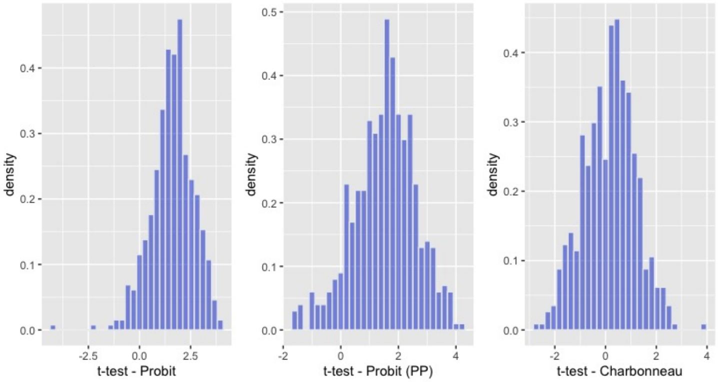


Figure 3: Histogram of the *t-test* for estimated $\beta^*_{21}$ in Design 7

For Charbonneau, *t-tests* are centered around zero, but not for Probit and Probit (PP). Biases of estimates for Probit and Probit (PP) are of the same magnitude $\rightarrow$ reflects mostly incidental parameter problem.

# Simulations: First stage QQ-plot for structural parameters

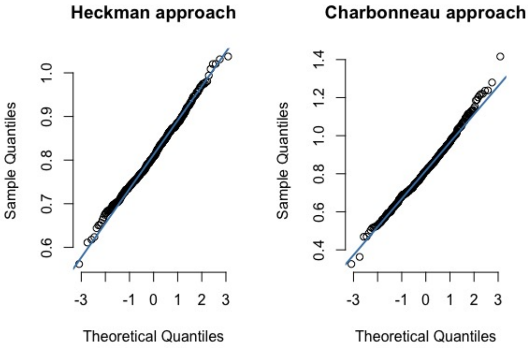


Figure 4: QQ plot of estimated $\beta^*_{21}$ in Design 6

All distributions are reasonably well approximated by a normal distribution.

# Simulations: First stage QQ-plot for structural parameters

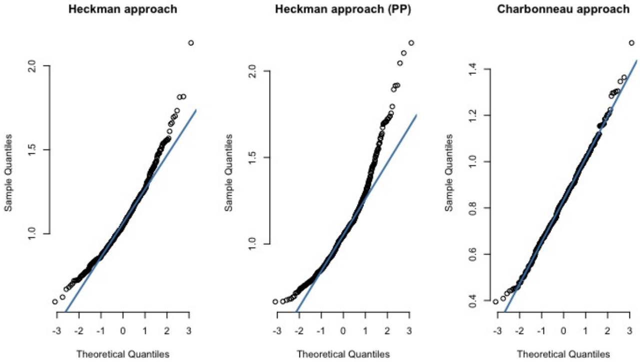


Figure 5: QQ plot of estimated $\beta^*_{21}$ in Design 5

# Simulations: First stage QQ-plot for structural parameters

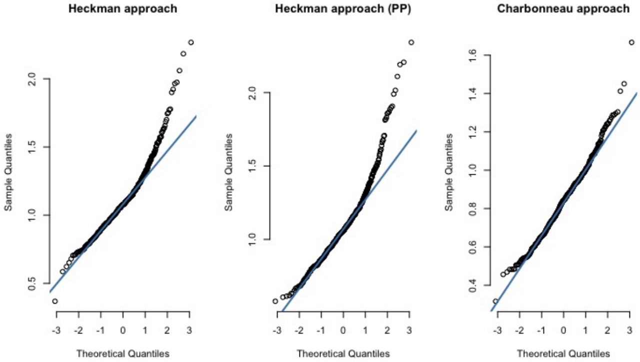


Figure 6: QQ plot of estimated $\beta^*_{21}$ in Design 7

In designs 5 and 7, Charbonneau are better approximated by a normal distribution. Other estimators are more skewed.

# Simulations: First stage fixed effects

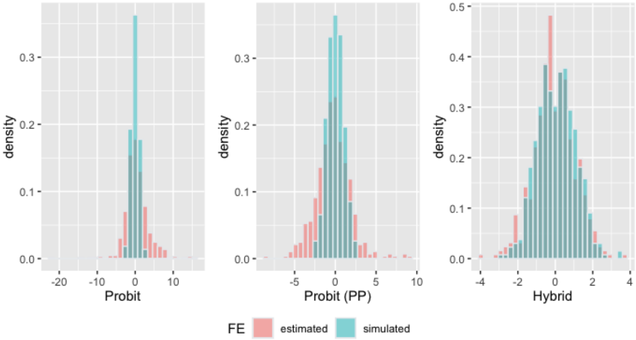


Figure 7: Histogram of estimated $\xi_1^*$ in Design 5

# Simulations: First stage fixed effects

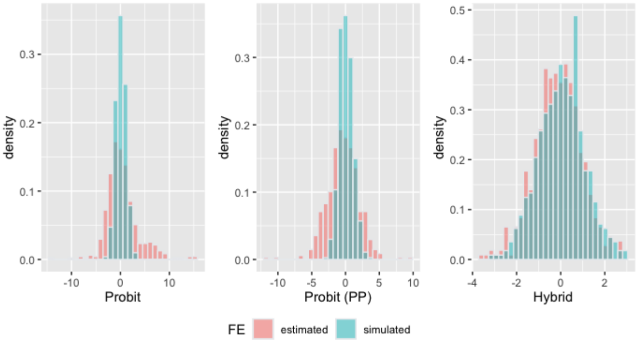


Figure 8: Histogram of estimated $\xi_1^*$ in Design 7

Probit estimates off due to: (i) not corrected for perfect prediction, and (ii) estimates are contaminated by the (asymptotic) bias of structural parameters.

Results of single indices

# Simulations: Second stage structural parameters mean bias

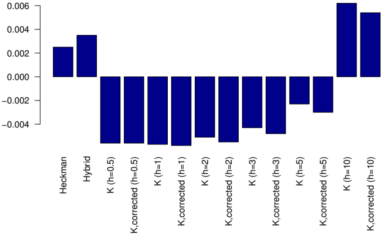


Figure 9: Mean bias of $\hat{\beta}_{11}$ in Design 6

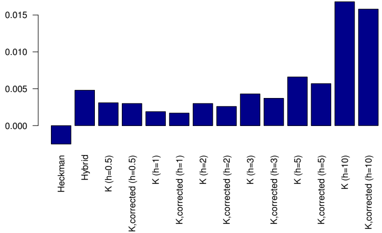


Figure 10: Mean bias of $\hat{\beta}_{12}$ in Design 6

For the modified Kyriazidou: $r = 1$ and $K$ set to be a standard normal density.

With no fixed effects, all proposed estimators have essentially no bias.

Even when including dummies in observation equations, estimators have nice properties.

# Simulations: Second stage structural parameters mean bias

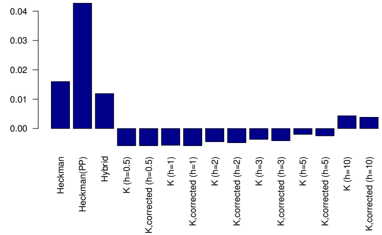


Figure 11: Mean bias of $\hat{\beta}_{11}$ in Design 5

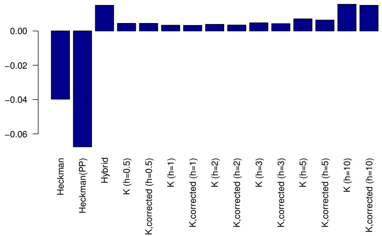


Figure 12: Mean bias of $\hat{\beta}_{12}$ in Design 5

Heckman and Heckman(PP) have biases: single indices are biased.

# Simulations: Second stage structural parameters mean bias

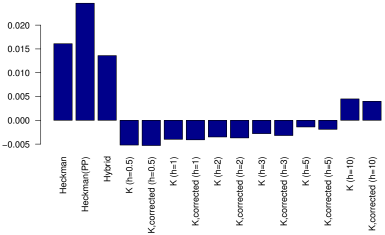


Figure 13: Mean bias of $\hat{\beta}_{11}$ in Design 7

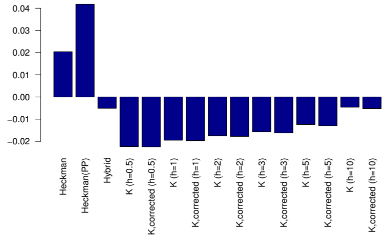


Figure 14: Mean bias of $\hat{\beta}_{12}$ in Design 7

Kyriazidou reduces further the bias for the continuous variable.

Hybrid reduces further for the binary variable.

Estimates sensitive to initial choice of $h$.

QQ plots

# Application to gravity model: Selection Equation

**Aim:** estimate how trade barriers affect both the decision of country $i$ to export to country $j$, and the volume of trade.

# Application to gravity model: Selection Equation

**Aim:** estimate how trade barriers affect both the decision of country $i$ to export to country $j$, and the volume of trade.

**Dataset:** provided by Helpman et al. [2008]. Information on directed trade flows and country characteristics for 158 countries in 1986.
Exclude Congo as exporter: avoid the problem of perfect prediction.

# Application to gravity model: Selection Equation

**Aim:** estimate how trade barriers affect both the decision of country $i$ to export to country $j$, and the volume of trade.

**Dataset:** provided by Helpman et al. [2008]. Information on directed trade flows and country characteristics for 158 countries in 1986.
Exclude Congo as exporter: avoid the problem of perfect prediction.

**First stage equation:** Same equation estimated by Helpman et al. [2008].

$$y_{2,ij} = \mathbb{1}(x'_{2,ij}\beta_2^* + \xi_i^* + \zeta_j^* > \eta_{ij}^*)$$

where:
(i) $y_{2,ij}$ is a binary variable, being one if the $i$ exports to $j$ and zero otherwise
(ii) $\xi_i^*$ is the exporter fixed effect
(iii) $\zeta_j^*$ is the importer fixed effect
(iv) $x_{2,ij}$ is the vector that collects the variables $Distance_{ij}$, $Common\ Border_{ij}$, $Colonial\ Ties_{ij}$, $Currency\ Union_{ij}$, $Common\ Legal\ System_{ij}$, $FTA_{ij}$, $Common\ Religion_{ij}$.

# Application to gravity model: Selection Equation

| | Probit | Charbonneau |
|---|---|---|
| *Common Language* | 0.2903*** (0.0379) | 0.4248*** (0.0732) |
| *Common Legal System* | 0.0972** (0.0296) | 0.1822*** (0.0597) |
| *Common Religion* | 0.2647*** (0.0585) | 0.4979*** (0.1125) |
| *Common Border* | −0.3798*** (0.0946) | −0.5164** (0.2300) |
| *Currency Union* | 0.4883*** (0.1306) | 1.0459*** (0.2362) |
| *Distance* | −0.6626*** (0.0208) | −1.0001*** (0.0541) |
| *FTA* | 2.0170*** (0.3085) | 3.5565*** (0.5237) |
| *Colonial Ties* | 0.3337 (0.2852) | 1.1432* (0.6473) |

Table 1: Estimates for the first stage. Standard errors are in parenthesis.
* indicates that the coefficient is significant at the 10 % level
** indicates that it is significant at the 5 % level
*** indicates that it is significant at the 1 % level.

# Application to gravity model: Observation Equation

**Second stage equation:** Difference to equation estimated by Helpman et al. [2008] is that we do not take into account $w_{ij}$.

$$y_{1,ij,t} = x'_{1,ij,t}\beta_1 + \vartheta_i + \chi_j + u_{ij,t}$$

where:
(i) $y_{1,ij}$ is the log of the value of the exports from $i$ to $j$
(ii) $\vartheta_i$ is the exporter fixed effect
(iii) $\chi_j$ is the importer fixed effect
(iv) $x_{1,ij,t}$ is the vector that collects the same variables as $x_{2,ij}$, with the exception of $Common\ Religion_{ij}$ (exclusion restriction)

**Equation estimated using only observations with positive trade flows $\rightarrow$ sample selection correction needed.**

# Application to gravity model: Observation Equation

| | Heckman (1) | Heckman (2) | Hybrid | K (h=0.5) | K (h=3) | K (h=10) |
|---|---|---|---|---|---|---|
| *Common Language* | 0.2090*** | 0.2333*** | 0.3566 | 0.1094 | 0.1830 | 0.1994 |
| *Legal System* | 0.4797*** | 0.4914*** | 0.4975 | 0.3805 | 0.4206 | 0.4255 |
| *Common Border* | 0.4031*** | 0.4318*** | 0.5836 | 0.6114 | 0.5673 | 0.5129 |
| *Currency Union* | 1.3537*** | 1.3368*** | 1.4456 | 1.2315 | 1.4164 | 1.4213 |
| *Distance* | −1.2087*** | −1.2190*** | -1.3470 | -1.1371 | -1.2503 | -1.1721 |
| *FTA* | 0.7774*** | 0.7719*** | 2.0569 | 1.2835 | 1.2408 | 0.2901 |
| *Colonial Ties* | 1.3418*** | 1.3434*** | 1.4209 | 1.2594 | 1.3387 | 1.1945 |
| *Common Religion* | 0.2389** | | | | | |
| *Inverse Mills-Ratio* | 0.2784*** | 0.2671*** | 0.7130 | | | |

Table 2: Estimates for the second stage equation. We denote by Kyriazidou by $K$.
* indicates that the coefficient is significant at the 10 % level
** indicates that the coefficient is significant at the 5 % level
*** indicates that the coefficient is significant at the 1 % level.

# Conclusion

- Accounting for sample selection in dyadic settings requires more involved methods than the standard Heckman [1979] approach.
- Difficulty arises through the incidental parameter problem in the selection equation.
- Charbonneau [2017] estimator:
  - Leads to estimates that are free of this problem in the first stage.
  - However, it does not deliver estimates of fixed effects essential to employ Heckman approach.
- To bypass this problem, we proposed:
  - To get the FE through a Hybrid estimation and further apply the Heckman procedure after a transformation is made in the selection equation.
    $\rightarrow$ it may not be well suited for sparse networks
  - A modification to the Kyriazidou [1997] estimator.
    $\rightarrow$ no need for FE estimates.
    $\rightarrow$ estimates are sensitive to chosen bandwidth.
    $\rightarrow$ needs a variable that satisfies exclusion restriction
- Drawbacks: computational costs!
- Simulations confirm the theoretical predictions that biases are reduced.

James E Anderson and Eric Van Wincoop. Gravity with gravitas: A solution to the border puzzle. *American economic review*, 93(1):170–192, 2003.

Karyne B Charbonneau. Multiple fixed effects in binary response panel data models. *The Econometrics Journal*, 20(3):S1–S13, 2017.

Iván Fernández-Val and Martin Weidner. Individual and time effects in nonlinear panel models with large n, t. *Journal of Econometrics*, 192(1):291–312, 2016.

James J Heckman. Sample selection as a specification error. *Econometrica*, 47: 153–161, 1979.

Elhanan Helpman, Marc Melitz, and Yona Rubinstein. Estimating trade flows: Trading partners and trading volumes. *The quarterly journal of economics*, 123 (2):441–487, 2008.

Ekaterini Kyriazidou. Estimation of a panel data sample selection model. *Econometrica: Journal of the Econometric Society*, pages 1335–1364, 1997.

Lung-Fei Lee. Generalized econometric models with selectivity. *Econometrica: Journal of the Econometric Society*, pages 507–512, 1983.

# References II


Jerzy Neyman and Elizabeth L Scott. Consistent estimates based on partially consistent observations. *Econometrica: Journal of the Econometric Society*, pages 1–32, 1948.

Jan Tinbergen. *Shaping the World Economy; Suggestions for an International Economic Policy*. Twentieth Century Fund, New York, January 1962. URL `http://hdl.handle.net/1765/16826`.

Jeffrey M Wooldridge, Do Won Kwak, and Robert S Martin. The robustness of conditional logit for binary response panel data models with serial correlation. *August. US Bureau of Labor Statistics Working Paper*, 502, 2018.

# Gravity model of Helpman et al. [2008] : Features

**Definition**

Multilateral resistance term: country's average trade barrier with all its trading partners (Anderson and Van Wincoop [2003]).

Two main features:

- Multilateral resistance terms.
- Theoretical framework that takes into account sample selection.
    - A fraction of firms in country $i$ decides to export to country $j$.
    - Firms have individual heterogeneous productivity and face variable and fixed costs when exporting.

# Gravity model of Helpman et al. [2008]

Considering that:

1. Country $i$ has $N_i$ heterogeneous firms.
2. Productivity is firm-specific given by $\frac{1}{a}$, $a$ follows a cumulative distribution function $G(a)$, with support $[a_L, a_H]$.
3. Firm in country $i$ has input costs $c_i a$ per unit of output.
4. If it sells in country $j$, it incurs:
    - Fixed costs $c_i f_{ij}$.
    - Variable costs (melting iceberg): $\tau_{ij}$ units of a product needs to be shipped from country $i$ to $j$ for one unit to arrive.
5. CES preferences.
6. Products are differentiated according to country of origin.
7. Monopolistic competition in the final products.
8. Economy of $I$ countries indexed by $i = 1, ...I$

# Gravity model of Helpman et al. [2008]

Model delivers the system of equations:

$$(1-\alpha)\left(\frac{\tau_{ij}c_i a_{ij}}{\alpha P_j}\right)^{1-\sigma} Y_j = c_i f_{ij} \tag{1}$$

where : $a_{ij}$ is defined as the point where profits are exactly zero, $P_j$ the price index of the economy of country $j$, $Y_j$ the size of its economy and $\sigma = 1/(1-\alpha)$ is the elasticity of substitution across products.

The fraction of country $i$'s $N_i$ firms that export to country $j$, is given by $G(a_{ij})$.

# Gravity model of Helpman et al. [2008]

$$V_{ij} = \begin{cases} \int_{a_L}^{a_{ij}} a^{1-\sigma} dG(a) & \text{for } a_{ij} \geq a_L \\ 0 & \text{otherwise} \end{cases} \tag{2}$$

where $V_{ij}$ determines the bilateral trade volume.

$$Y_{1,ij} = \left( \frac{c_i \tau_{ij}}{\alpha P_j} \right)^{1-\sigma} Y_j N_i V_{ij} \tag{3}$$

defines the value of country $j$'s imports from $i$ ($Y_{1,ij}$).

$$P_j^{1-\sigma} = \sum_{i=1}^{I} \left( \frac{c_i \tau_{ij}}{\alpha} \right)^{1-\sigma} N_i V_{ij} \tag{4}$$

defines the price indices of country $j$.

# Estimation method of Helpman et al. [2008]

## Assumption 1:

Firm productivity $\frac{1}{a}$ is Pareto distributed, with support $[a_L, a_H]$.

## Assumption 2:

There are i.i.d. unmeasured country-pair specific trade frictions $u_{ij,t} \sim N(0, \sigma_u^2)$ such that: $\tau_{ij,t}^{\sigma-1} \equiv D_{ij,t}^{\gamma} e^{-u_{ij,t}}$
where $D_{ij,t}$ is the symmetric distance between $i$ and $j$.

## Assumption 3:

There are i.i.d. unmeasured country-pair specific trade frictions $\nu_{ij,t} \sim N(0, \sigma_\nu^2)$ that may be correlated with $u_{ij,t}$, such that:
$f_{ij,t} \equiv \exp\left(\phi_{EX,i} + \phi_{IM,j} + \kappa\phi_{ij,t} - \nu_{ij,t}\right)$
where $\phi_{IM,j}$ is a fixed trade barrier imposed by the importing country, $\phi_{EX,i}$ is a measure of fixed export costs, and $\phi_{ij,t}$ is an observed measure of country-pair specific fixed trade costs.

# Estimation method of Helpman et al. [2008]

Given Assumptions 1, 2 and 3, and by log-linearizing the gravity equation, which defines the export volume from country $i$ to $j$:

$$y_{1,ij,t} = x'_{1,ij,t}\beta_1 + \vartheta_i + \chi_j + w_{ij,t} + u_{ij,t}, \tag{5}$$

where:
(i) $y_{1,ij,t} = \ln Y_{1,ij,t}$
(ii) $w_{ij,t} = \ln W_{ij,t}$
(iii) $\beta_1$ is a vector that collects the coefficients of the remaining structural parameters, $\beta_1 = (\beta_0, \gamma_1')'$
(iv) $x_{1,ij,t}$ is a vector that collects 1 and the vector $d_{ij,t} = \ln D_{ij,t}$
(v) $\chi_j = (\sigma - 1)\ln P_j + \ln Y_j$, which is an importer fixed effects
(vi) $\vartheta_i = -(\sigma - 1)\ln c_i + \ln N_i$, which is an exporter fixed effects.

Important difference to the equation derived in previous studies: the new variable $w_{ij,t}$, which controls for the fraction of firms exporting.

# Estimation method of Helpman et al. [2008]

Latent variable $Y^*_{2,ij,t}$ is given by the ratio of variable export profits for the most productive firm (with productivity $1/a_L$) to the fixed export costs for exports from $i$ to $j$ (expressions that are given by the zero profit condition). Then, in this case, positive exports are observed if $Y^*_{2,ij,t} > 1$.

By log-linearizing the equation for the latent variable $Y^*_{2,ij,t}$, and given Assumption 3, we have that:

$$y^*_{2,ij,t} = x'_{2,ij,t}\beta_2 + \xi_i + \zeta_j + \eta_{ij,t}, \tag{6}$$

where:
(i) $\eta_{ij,t} \equiv u_{ij,t} + \nu_{ij,t} \sim N(0, \sigma^2_u + \sigma^2_\nu)$ is i.i.d., but correlated with $u_{ij,t}$
(ii) $\xi_i = -\sigma \ln c_i + \phi_{EX,i}$ is an exporter fixed effect
(iii) $\zeta_j = (\sigma - 1) \ln P_j + \ln Y_j - \phi_{IM,j}$ is an importer fixed effect
(iv) $\beta_2$ is a vector that collects the coefficients of the remaining structural parameters, $\beta_2 = (\gamma_0, \gamma'_2)'$
(iv) $x_{2,ij,t}$ is a vector that collects 1, and the vectors $d_{ij,t} = \ln D_{ij,t}$, and $\phi_{ij,t}$.

$x_{2,ij,t}$ includes both the regressors in $x_{1,ij,t}$ and additional regressors that affect only the decision to export, but not the volume of exports (exclusion restriction).

# The Heckman's steps

The least squares estimators of $\beta_1$ and $\sigma_{u\eta^*}$ are unbiased but inefficient, due to the heteroskedasticity in $\mathbb{E}[\nu^2_{ij,t}]$, as shown in Heckman [1979]:

$$\mathbb{E}[\nu^2_{ij,t}] = \sigma^2_u\Big((1-\rho^2) + \rho^2(1 + z_{ij,t}\lambda_{ij,t} - \lambda^2_{ij,t})\Big)$$

- Step 1: Estimate the probability that $y^{**}_{2,ij,t} \geq 0$ using probit analysis on the sample, given by the selection equation.
- Step 2: From this estimator (provided that it is consistent), one can obtain $\hat{z}_{ij,t}$ consistently.
- Step 3: The estimated value of $\lambda_{ij,t}(z_{ij,t})$ is used as a regressor in the observation equation fit on the subsample. The regression estimators are then consistent for $\beta_1$ and $\sigma_{u\eta^*}$.
- Step 4: One can then consistently estimate $\sigma^2_u$ from the following. From step (3), we consistently estimate $\sigma_{u\eta^*}$, through the estimator $\hat{\sigma}_{u\eta^*}$. Denote the residuals for each observation from step 3 as $\hat{\nu}_{ij,t}$. Then, using $\hat{z}_{ij,t}$ and $\hat{\lambda}_{ij,t}$ the estimated values from step (2), an estimator of $\sigma^2_u$ is:

$$\hat{\sigma}^2_u = \frac{\sum_{i=1}^{N_i}\sum_{j\neq i}\sum_{t=1}^{T}\hat{\nu}^2_{ij,t}}{N_i(N_i-1)T} - \frac{\hat{\sigma}_{u\eta^*}}{N_i(N_i-1)T}\sum_{i=1}^{N_i}\sum_{j\neq i}\sum_{t=1}^{T}(\hat{\lambda}_{ij,t}\hat{z}_{ij,t} - \hat{\lambda}^2_{ij,t})$$

# Incidental parameter problem

Heckman [1979]'s approach:

- **Stage 1:** Estimate the selection equation by MLE (probit).
- **Stage 2:** Obtain inverse Mills-ratio and estimate the observation equation by FGLS.

**But** fixed effects estimators in nonlinear models suffer from the **incidental parameter problem** (Neyman and Scott [1948]).

**How does this problem arise?**
$y_{2,ij,t}$ is generated by the process:

$$y_{2,ij,t}|x_{2,ij,t}, \xi^*, \zeta^*, \beta_2^* \sim f_{Y_2}(\cdot|x_{2,ij,t}, \xi^*, \zeta^*, \beta_2^*)$$

where: $\xi^* = (\xi_1^*, ...\xi_N^*)$, $\zeta^* = (\zeta_1^*, ...\zeta_N^*)$, $f_{Y_2}$ is a known probability function and $\xi_i^*, \zeta_j^*$ are the unobserved fixed effects.

# Incidental parameter problem

Using a single-index specification with fixed effects:

$$f_{Y_2}(y_{2,ij,t}|x_{2,ij,t},\xi^*,\zeta^*,\beta_2^*) = \Phi(x_{2,ij,t}'\beta_2^* + \xi_i^* + \zeta_j^*)^{y_{2,ij,t}} \times [1-\Phi(x_{2,ij,t}'\beta_2^* + \xi_i^* + \zeta_j^*)]^{1-y_{2,ij,t}}, \quad (7)$$

To estimate the parameters, we solve the sample analogue of:

$$\max_{(\beta_2^*,\omega_{NN}^*)\in\mathbb{R}^{\dim\beta_2^*+\dim\omega_{NN}^*}} \mathcal{L}(\beta_2^*,\omega_{NN}^*) \quad (8)$$

with $\omega_{NN}^* = (\xi_1^*,...\xi_N^*,\zeta_1^*,...\zeta_N^*)'$, and:

$$\mathcal{L}(\beta_2^*,\omega_{NN}^*) \quad (9)$$
$$= (N(N-1)T)^{-1}\left\{\sum_{i=1}^{N}\sum_{j\neq i}\sum_{t=1}^{T}\log f_{Y_2}(y_{2,ij,t}|x_{2,ij,t},\xi^*,\zeta^*,\beta_2^*) - b(\iota_{NN}'\omega_{NN}^*)^2/2\right\}$$

$b>0$ is an arbitrary constant, $\iota_{NN} = (1_N', -1_N')'$ and $1_N$ denotes a vector of ones of dimension $N$.

# Incidental parameter problem

For a given $\beta_2^*$, the optimal $\hat{\omega}_{NN}^*$ is:

$$\hat{\omega}_{NN}^*(\beta_2^*) = \underset{\omega_{NN}^* \in \mathbb{R}^{\dim \omega_{NN}^*}}{\arg\max} \mathcal{L}(\beta_2^*, \omega_{NN}^*) \tag{10}$$

Then, the fixed effects estimator of $\beta_2^*$ and $\omega_{NN}^*$ are:

$$\hat{\beta}_2^* = \underset{\beta_2^* \in \mathbb{R}^{\dim \beta_2^*}}{\arg\max} \mathcal{L}(\beta_2^*, \hat{\omega}_{NN}^*(\beta_2^*)) \tag{11}$$

$$\hat{\omega}_{NN}^*(\beta_2^*) = \hat{\omega}_{NN}^*(\hat{\beta}_2^*) \tag{12}$$

**The source of the problem is that the dimension of the nuisance parameters increase with sample size and their estimation converge at a slower rate than the structural parameters.**

# Incidental parameter problem

Denote:

$$\bar{\beta}_2^* = \underset{\beta_2^* \in \mathbb{R}^{\dim \beta_2^*}}{\arg\max}\ \mathbb{E}_\omega \Big[ \mathcal{L}(\beta_2^*, \hat{\omega}_{NN}^*(\beta_2^*)) \Big]$$

Using an asymptotic expansion for smooth likelihoods under appropriate regularity conditions (Fernández-Val and Weidner [2016]):

$$\bar{\beta}_2^* = \beta_{2,0}^* + \frac{\bar{B}_\infty}{(N-1)T} + \frac{\bar{D}_\infty}{(N-1)T} + o_P(((N-1)T)^{-1}) \tag{13}$$

For some constants $\bar{B}_\infty$ and $\bar{D}_\infty$. By the properties of the maximum likelihood estimator:

$$\sqrt{N(N-1)T}(\hat{\beta}_2^* - \bar{\beta}_2^*) \xrightarrow{d} N(0, \bar{V}_{B\infty}) \tag{14}$$

For some $\bar{V}_{B\infty}$. By Slutsky Theorem:

$$\sqrt{N(N-1)T}(\hat{\beta}_2^* - \beta_{2,0}^*) \xrightarrow{d} N\Big(\frac{\bar{B}_\infty}{\sqrt{T}} + \frac{\bar{D}_\infty}{\sqrt{T}}, \bar{V}_{B\infty}\Big)$$

# Incidental parameter problem: asymptotic expansion

Taking a first-order Taylor expansion of the first order conditions of Equation (11) around $\beta^*_{2,0}$, gives:

$$0 = \frac{\partial \mathcal{L}(\hat{\beta}^*_2, \hat{\omega}^*_{NN}(\beta^*_2))}{\partial_{\beta^*_2}} \approx \frac{\partial \mathcal{L}(\beta^*_{2,0}, \hat{\omega}^*_{NN}(\beta^*_{2,0}))}{\partial_{\beta^*_2}} - \bar{W}_\infty \sqrt{N(N-1)T}(\hat{\beta}^*_2 - \beta^*_{2,0}) \tag{15}$$

Then, we apply a second-order Taylor expansion to approximate the above term $\frac{\partial \mathcal{L}(\beta^*_{2,0}, \hat{\omega}^*_{NN}(\beta^*_{2,0}))}{\partial_{\beta^*_2}}$ around $\omega^*_{NN}(\beta^*_{2,0})$, such that the estimates of the fixed effects are taken into account.

$$\frac{\partial \mathcal{L}(\beta^*_{2,0}, \hat{\omega}^*_{NN}(\beta^*_{2,0}))}{\partial_{\beta^*_2}} \approx \frac{\partial \mathcal{L}(\beta^*_{2,0}, \omega^*_{NN}(\beta^*_{2,0}))}{\partial_{\beta^*_2}} \tag{16}$$

$$+ \frac{\partial^2 \mathcal{L}(\beta^*_{2,0}, \omega^*_{NN}(\beta^*_{2,0}))}{\partial_{\beta^*_2}\partial_{\omega'_{NN}}}[\hat{\omega}^*_{NN}(\beta^*_{2,0}) - \omega^*_{NN}(\beta^*_{2,0})]$$

$$+ \sum_{k=1}^{\dim \omega_{NN}} \frac{\partial^3 \mathcal{L}(\beta^*_{2,0}, \omega^*_{NN}(\beta^*_{2,0}))}{\partial_{\beta^*_2}\partial_{\omega'_{NN}}\partial_{\omega_{NN,k}}}[\hat{\omega}^*_{NN}(\beta^*_{2,0}) - \omega^*_{NN}(\beta^*_{2,0})][\hat{\omega}^*_{NN,k}(\beta^*_{2,0}) - \omega^*_{NN,k}(\beta^*_{2,0})]/2$$

# Incidental parameter problem: asymptotic expansion

Under regularity conditions, since the first term in this expression is the score vector, it has mean zero and it generates the asymptotic variance. By the information matrix equality and the Central Limit Theorem, we have:

$$\frac{\partial \mathcal{L}(\beta^*_{2,0}, \omega^*_{NN}(\beta^*_{2,0}))}{\partial_{\beta^*_2}} \xrightarrow{d} N(0, \bar{W}_\infty) \tag{17}$$

For some variance $\bar{W}_\infty$. According Fernández-Val and Weidner [2016], the second and the third term satisfies:

$$\begin{aligned}
&\frac{\partial^2 \mathcal{L}(\beta^*_{2,0}, \omega^*_{NN}(\beta^*_{2,0}))}{\partial_{\beta^*_2}\partial_{\omega'_{NN}}}[\hat{\omega}^*_{NN}(\beta^*_{2,0}) - \omega^*_{NN}(\beta^*_{2,0})] \\
&+ \sum_{k=1}^{\dim \omega_{NN}} \frac{\partial^3 \mathcal{L}(\beta^*_{2,0}, \omega^*_{NN}(\beta^*_{2,0}))}{\partial_{\beta^*_2}\partial_{\omega'_{NN}}\partial_{\omega_{NN,k}}}[\hat{\omega}^*_{NN}(\beta^*_{2,0}) - \omega^*_{NN}(\beta^*_{2,0})][\hat{\omega}^*_{NN,k}(\beta^*_{2,0}) - \omega^*_{NN,k}(\beta^*_{2,0})]/2 \\
&\approx \sqrt{N(N-1)T}\Big(\frac{\bar{B}^\beta_\infty}{(N-1)T} + \frac{\bar{D}^\beta_\infty}{(N-1)T}\Big)
\end{aligned} \tag{18}$$

# Incidental parameter problem: asymptotic expansion

The analytical form of terms $\bar{B}_{\infty}^{\beta}$ and $\bar{D}_{\infty}^{\beta}$ can be obtained from the second-order Taylor expansion as shown in Fernández-Val and Weidner [2016]. Those terms originate from elements corresponding to the two-way fixed effects.
Plugging in the expression (18) into the equation for the first-order Taylor expansion, we have, as $N \to \infty$:

$$\bar{W}_{\infty}\sqrt{N(N-1)T}(\hat{\beta}_2^* - \beta_{2,0}^*) = \frac{\partial \mathcal{L}(\beta_{2,0}^*, \omega_{NN}^*(\beta_{2,0}^*))}{\partial_{\beta_2^*}} + \frac{\bar{B}_{\infty}^{\beta}}{\sqrt{T}} + \frac{\bar{D}_{\infty}^{\beta}}{\sqrt{T}} \tag{19}$$

By Slutsky Theorem, we have, given (17):

$$\sqrt{N(N-1)T}(\hat{\beta}_2^* - \beta_{2,0}^*) \xrightarrow{d} \bar{W}_{\infty}^{-1} N\Big(\frac{\bar{B}_{\infty}^{\beta}}{\sqrt{T}} + \frac{\bar{D}_{\infty}^{\beta}}{\sqrt{T}}, \bar{W}_{\infty}\Big) \tag{20}$$

Therefore, compared to the expression given in the presentation, we have that:

$$\bar{W}_{\infty}^{-1}\bar{B}_{\infty}^{\beta} = \bar{B}_{\infty}$$

$$\bar{W}_{\infty}^{-1}\bar{D}_{\infty}^{\beta} = \bar{D}_{\infty}$$

# Conditional logit estimation: Estimator

Denote:

- $m_n$ distinct quadruples in the dataset.
- a function $\sigma$ that maps the possible quadruples to an index set $N_{m_n} = \{1, 2, ..., m_n\}$.

Define the variables:

$$z(\sigma\{l, i; j, k\}) = \frac{(y_{2,lj} - y_{2,lk}) - (y_{2,ij} - y_{2,ik})}{2}$$
$$r(\sigma\{l, i; j, k\}) = (x_{2,lj} - x_{2,lk}) - (x_{2,ij} - x_{2,ik}))$$

Event that $z \in \{-1, 1\}$ corresponds to the condition that for any $ij$, $l$ and $k$ satisfies $y_{2,lk} + y_{2,lj} = 1, y_{2,ij} + y_{2,ik} = 1, y_{2,ij} + y_{2,lj} = 1$.

**Conditional on $x_2$ and $z \in \{-1, 1\}$, the distribution of $z$ is logistic and does not depend on the fixed effects.**

- When $z = 1$, we have necessarily that $y_{2,lk} = 1$, and when $z = -1$, it is necessarily zero.

# Conditional logit estimation: Estimator

The estimator is given by:

$$\hat{\beta}_2^* = \arg\max_{\beta_2^* \in \Theta} \mathcal{L}_{CL}(\beta_2^*)$$

$\Theta$ refers to the parameter space searched over, and $\mathcal{L}_{CL}$ is the objective function given by:

$$\begin{aligned}
\mathcal{L}_{CL}(\beta_2^*) = \sum_{\sigma \in N_{m_n}} & \mathbb{1}\{z(\sigma\{l,i;j,k\}) = 1\} \log(F(r(\sigma\{l,i;j,k\})'\beta_2^*)) \\
& + \mathbb{1}\{z(\sigma\{l,i;j,k\}) = -1\} \log(1 - F(r(\sigma\{l,i;j,k\})'\beta_2^*))
\end{aligned}$$

Denote:

- $m_n^*$ the number of quadruples that contributes to the likelihood.
- $p_n$ the expected fraction of such quadruples over the total quadruples.

# Conditional logit estimation: Asymptotic Properties

## Assumption 9:

$\beta^*_{2,0}$ is interior to $\Theta$, which is a compact subset of $\mathbb{R}^{\dim \beta^*_2}$.

## Assumption 10:

For all $(i, j) \in N \times N, \mathbb{E}(||x_{2,ij}||^2) < c$, where $c$ is a finite constant.

## Assumption 11:

$Np_n \to \infty$ as $N \to \infty$ and the matrix

$$\lim_{N \to \infty} (m_n p_n)^{-1} \sum_{\sigma \in N_{m_n}} \mathbb{E}(r(\sigma\{l, i; j, k\}) r(\sigma\{l, i; j, k\})' f(r(\sigma\{l, i; j, k\})' \beta^*_{2,0}) \mathbb{1}\{z \in \{-1$$

has maximal rank, where $f$ is the density of the logistic function.

## Theorem 2:

Let Assumptions 8 - 11 hold. Then, $\hat{\beta}^*_2 \xrightarrow{p} \beta^*_{2,0}$ as $N \to \infty$.

# Conditional logit estimation: Asymptotic Properties

To establish the limiting distribution, a stronger form of Assumption 10 is made, Then, as $N \to \infty$:

**Theorem 3:**

Let Assumptions 8 - 12 hold. Then $||\hat{\beta}_2^* - \beta_{2,0}^*|| = O_p(1/\sqrt{N(N-1)p_n})$ and

$$\Omega^{-1/2}(\hat{\beta}_2^* - \beta_{2,0}^*) \xrightarrow{d} N(0, I)$$

where:

$$s(\sigma; \beta_2^*) = r_\sigma \{\mathbb{1}\{z_\sigma = 1\}(1 - F(r_\sigma{}'\beta_2^*)) - \mathbb{1}\{z_\sigma = -1\}F(r_\sigma{}'\beta_2^*)\},$$

$$\upsilon_{ij}(\beta_2^*) = \sum_{i \neq l, j} \sum_{k \neq i, l, j} s(\sigma\{l, i; j, k\}; \beta_2^*) \qquad \Upsilon(\beta_2^*) = \sum_{l=1}^{N} \sum_{j \neq l} \upsilon_{ij}\upsilon_{ij}'$$

$$H(\beta_2^*) = \frac{\partial^2 \mathcal{L}_{CL}(\beta_2^*)}{\partial \beta_2^* \partial \beta_2^*{}'} = \sum_{\sigma \in N_{m_n}} r_\sigma r_\sigma' f(r_\sigma' \beta_2^*) \mathbb{1}\{z \in \{-1, 1\}\}$$

$$\Omega = H(\hat{\beta}_2^*)^{-1} \Upsilon(\hat{\beta}_2^*) H(\hat{\beta}_2^*)^{-1}$$

# Second approach: Choice of bandwidth

For a given order of differentiability $r$ of the expression $\mathbb{E}(\Delta x_1' \varphi \Phi | \Delta x_2' \beta_2^*)$ and a given sample size $N(N-1)$, $h_n = h(N(N-1))^{-\mu}$ should be chosen such that $\mu = 1/(2(r+1)+1)$.

Value for $\mu$ comes from:

- Rate of convergence of the distribution of $\hat{\beta}_1$ is maximized by setting $\mu$ as small as possible.
- $\mu$ should be in the range $1 - 2p < \mu < p/2$, where $p$ is the rate of convergence of $\hat{\beta}_2^*$.
- $\hat{\beta}_2^*$ should converge fast enough: at least at a rate $p > (r+1)/(2(r+1)+1)$.

**Problem of choosing a bandwith boils down to choosing a constant $h$.**
Kyriazidou [1997]: for any positive initially chosen $h$, the distance between the estimator and the parameters' true values is minimized through a correction.

# Second approach: Choice of bandwidth

**Corollary 1:**

Let $\hat{\beta}_1$ be the estimator with window width $h_n = h(N(N-1))^{-1/(2(r+1)+1)}$, and $\hat{\beta}_{1,\delta}$ the estimator with window width $h_{n,\delta} = h(N(N-1))^{-\delta/(2(r+1)+1)}$ where $\delta \in (0,1)$.
Define:

$$\hat{\hat{\beta}}_1 = \frac{\hat{\beta}_1 - N(N-1)^{-(1-\delta)(r+1)/(2(r+1)+1)}\hat{\beta}_{1,\delta}}{1 - N(N-1)^{-(1-\delta)(r+1)/(2(r+1)+1)}}$$

Then, $(\hat{\hat{\beta}}_1 - \beta_1)$ will converge to a normal distribution centered around 0 at rate $(N(N-1))^{-(r+1)/(2(r+1)-1)}$.

**Given an appropriately chosen $h_n$, using an estimated value for $\beta_2^*$ does not affect the limiting distribution of the estimator for $\beta_1$.**

# Simulation: single indices

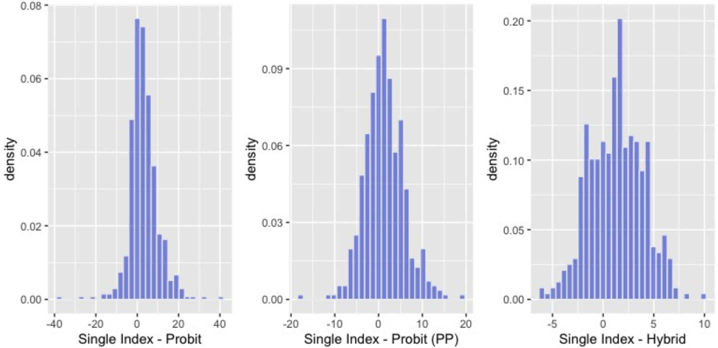


Figure 15: Histogram of $\hat{z}_{12} - z_{12}$ in Design 5

# Simulation: single indices

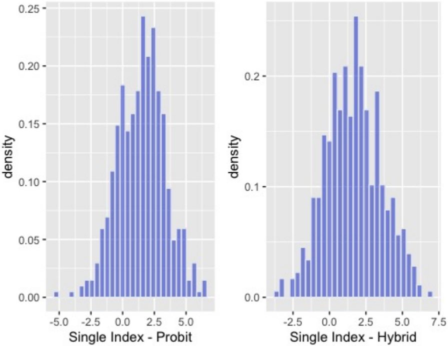


Figure 16: Histogram of $\hat{z}_{12} - z_{12}$ in Design 6

# Simulation: single indices

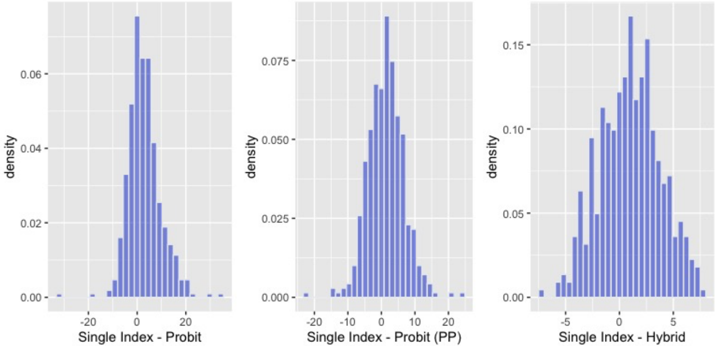


Figure 17: Histogram of $\hat{z}_{12} - z_{12}$ in Design 7

# Simulations: First stage QQ plot for single indices

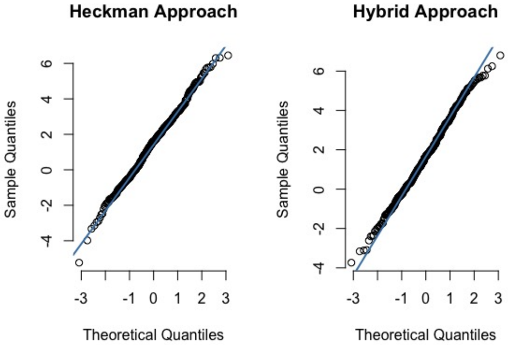


Figure 18: QQ plot of $\hat{z}_{12} - z_{12}$ in Design 6

With no fixed effects, the distributions are overall well approximated by a normal distribution.

# Simulations: First stage QQ plot for single indices

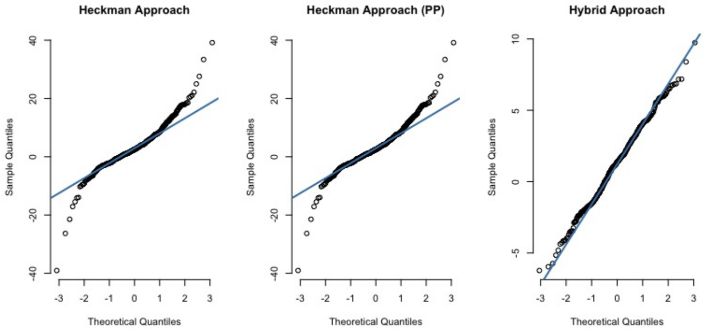


Figure 19: QQ plot of $\hat{z}_{12} - z_{12}$ in Design 5

In Design 5, the distributions for the Probit and Probit(PP) have heavier tails than the normal distribution $\rightarrow$ mainly affected by the extreme value of fixed effects.

# Simulations: First stage QQ plot for single indices

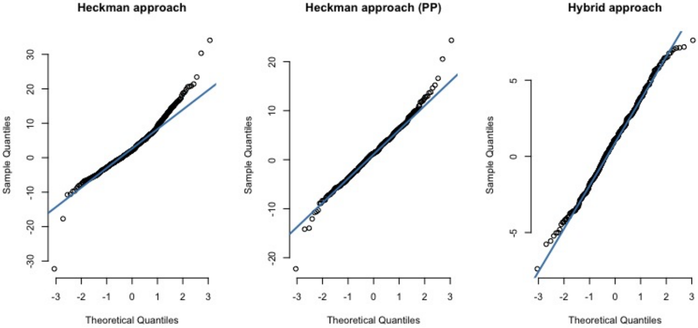


Figure 20: QQ plot of $\hat{z}_{12} - z_{12}$ in Design 7

In Design 7, the distributions for the Probit and Probit(PP) are more skewed $\rightarrow$ mainly affected by the distribution of the structural parameters.

# Simulation: observation equation

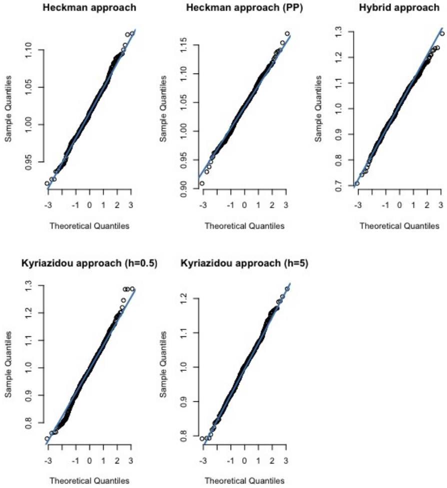


Figure 21: QQ plot of estimated $\beta_{11}$ in Design 5

# Simulation: observation equation

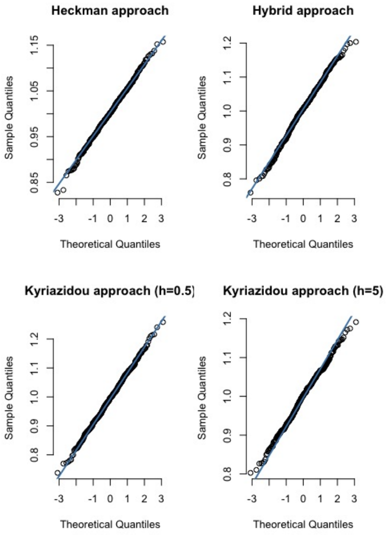


Figure 22: QQ plot of estimated $\beta_{11}$ in Design 6

# Simulation: observation equation

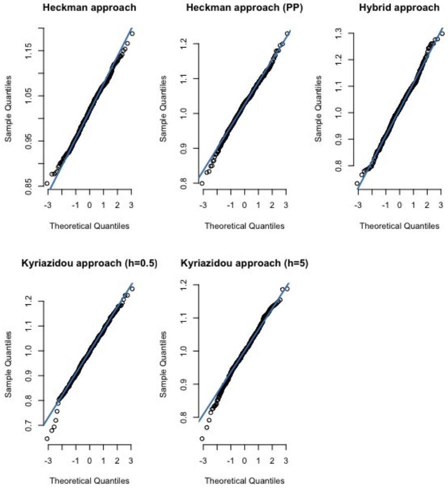


Figure 23: QQ plot of estimated $\beta_{11}$ in Design 7