Estimating Average Partial Effects of Fixed Effect Panel Logit Models

Laura Liu Alexandre Poirier Ji-Liang Shiu

National Chung Cheng University

March 23, 2020

Panel Data Models

- A panel data $\{y_t, x_{t1}, x_{t2}, ..., x_{tp} : t = 1, ..., T\}$ is at least two time periods are available for all cross section observations.
- We are interested in unobserved effects panel data models (Motivation: The Omitted Variables Problem).
- An unobserved, time-constant variable c is called an unobserved effect in panel data analysis.
- The partial effects of the observable explanatory variables $x_{tj} : j = 1, ...p$ in the population regression function

$$Y_t = X_t\beta_0 + C + U_t$$

- We can difference across the two time periods to eliminate the time-constant unobservable C.
  Define $\triangle Y = Y_2 - Y_1 \quad \triangle U = U_2 - U_1 \quad \triangle X = X_2 - X_1$
  $\Rightarrow \triangle Y = \triangle X\beta_0 + \triangle U$, just a standard linear model

Binary Panel Data Models

- The linear probability model (LPM) for binary response $Y_t \in \{0, 1\}$ is specified as

$$Y_t = X_t\beta_0 + C + U_t,$$

  where X_t is a $1\times p$ row vector of covariates, β_0 is a $p\times 1$ column vector of coefficients, C is the unobserved individual effect, and U_t is a continuously distributed disturbance independent of X_t and C.
- $P(Y_t = 1|X_t, C) = X_t\beta_0 + C$: Interpretation of β_{01}: the change in the probability of success given a one-unit increase in X_{01}
- Model mis-specification: predicted probabilities may bigger than one or less than zero

Short Panel Data Logit Models

- Consider binary panel data logit models and the models can be derived from an underlying latent variable model

$$Y_t = 1(\underbrace{X_t\beta_0 + C + E_t}_{\text{latent variable formulation}} > 0),$$

  where X_t is a $1 \times p$ row vector of covariates, β_0 is a $p \times 1$ column vector of coefficients, C is the unobserved individual effect, E_t is a continuously distributed disturbance independent of X_t and C, and E_t has a standard logistic distribution.
- The response probability is

$$P(Y_t = 1|X_t, C) = \frac{\exp(X_t\beta_0 + C)}{1 + \exp(X_t\beta_0 + C)} = \Lambda(X_t\beta_0 + C).$$

- Unfortunately, in addition to being computationally difficult, estimation of the C (fixed effect;parameter) along with β_0 introduces an incidental parameters problem for MLE.

Average Partial Effects

- Define the average structural function (ASF) by integrating out the response probability across the marginal distribution of the unobserved effect C.

$$\begin{aligned}\mathsf{ASF}(x_t) &= \int_{\mathcal{C}} \mathsf{P}\left(Y_t = 1 | X_t = x_t, C = c\right) f_C(c) dc \\ &= \int_{\mathcal{C}} \Lambda(x_t\beta_0 + c) f_C(c) dc.\end{aligned}$$

- The average partial effect (APE) is defined by taking derivatives of ASF with respect to continuous elements of x_t, or differences with respect to discrete elements of x_t.

$$\mathsf{APE}(x_{tk}) = \beta_{0k} \int_{\mathcal{C}} \frac{\partial \Lambda(x_t\beta_0 + c)}{\partial x_{tk}} f_C(c) dc.$$

- Estimating ASFs and APEs requires specifying or knowing a population distribution of C.

Empirical Examples in the Existing Literature

- (**Panel Binary-choice Logit Model with an Unobserved Effects**) The corresponding conditional distribution of interest:

$$f_{Y_t|X_t,C} = \Lambda(X_t\beta_0 + C)^{Y_t} \times (1 - \Lambda(X_t\beta_0 + C))^{1-Y_t},$$

  where Λ is the CDF of the random logistic shock E_t.
- Empirical applications:
  1. Health status movement
  2. Welfare participation
  3. Labor force participation of married women

Panel Data Studies

- In the panel data literature, there are two approaches to tackling the unobserved heterogeneity C_i: fixed effects and random effects.
- **Linear probability models (LPM)** with an additive unobserved effect: applying an appropriate transformation to eliminate the unobserved effect and then implementing instrument variables (IV) in a GMM framework. Anderson and Hsiao (1982), Arellano and Bond (1991), Arellano and Bover (1995) and Ahn and Schmidt (1995).
- **Fixed Effect:** Utilizing the conditional probability approach for certain types of discrete choice logit models to estimate the unknown parameters. Chamberlain (1984, Static), Honoré and Kyriazidou (2000, Dynamic).
- Chamberlain (2010) considered the identification of binary response models of short panel and showed that identification of the parameter is possible only in the logistic case.

Related Studies

- **Nonlinear Models**: Arellano and Carrasco (2003), Altonji and Matzkin (2005), Hoderlein and Mammen (2007), Bester and Hansen (2009), Chen and Swanson (2012), Hoderlein and White (2012), Graham and Powell (2012), Hoderlein and White (2012), Browning and Carro (2014), Chernozhukov, Fernández-Val, Hahn, and Newey (2013), and Chernozhukov, Fernandez-Val, Hoderlein, Holzmann, and Newey (2012).
- **Random Effects**:
  Wooldridge (2005), Alvarez and Arellano (2003), Arellano and Bonhomme (2009)
- Honoré and Lewbel (2002) provided a set of conditions for identification of the parameters of a binary choice model allowing for general predetermined explanatory variables and propose a root-n consistent GMM estimator to estimate the parameters.

Panel Data Studies

- Random effect approaches are based on an average likelihood:

$$f_{Y_t|X_t}(y_t|x_t;\beta,\theta) = \int f_{Y_t|X_t,C}(y_t|x_t,c;\beta) f_{C|X_t}(c|x_t;\theta)dc.$$

  - **Random Effects**: Restriction $D(C|X)$: independence and *normal*
  - **Correlated Random Effects**: Wooldridge (2005), suggests a general method for handling the initial conditions problem by using *a density function of unobserved heterogeneity conditional on the strictly exogenous variables and the initial condition*. Restriction $D(C|X)$: linear mean and *normal*

Features of Methods for Binary Response Models

Methods	$P(y_{it} = 1 \vert X_{it}, C_i)$ Bounded in $(0,1)$?	Restricts $E(C_i \vert X_i)$?	APEs?
LPM, within	No	No	Yes
RE probit	Yes	indep.,normal	Yes
CRE probit	Yes	linear mean,normal	Yes
FE logit	Yes	No	No

Motivation & Contributions

- 1. Propose an estimation method for the APE of the fixed effect logit model (only identify the parameter)
  2. Chances of misspecification increase as we need to specify $f_{C|X_t}(c|x_t;\theta)$ as normally distributed; identify a distribution for C without a normality assumption.
- An average likelihood can be constructed as follows:

$$f(y_t|x_t,\overline{w}) = \int \underbrace{f_{Y_t|X_t,C}(y_t|x_t,c;\beta_0)}_{\text{linear index structure}} \cdot \underbrace{f_{C|\overline{W}}(c|\overline{w})}_{\text{require for APE}} \, dc,$$

where $f_{Y_t|X_t,C} = \Lambda(X_t\beta_0 + C)^{Y_t} \times (1 - \Lambda(X_t\beta_0 + C))^{1-Y_t}$.

- Provide the internally consistent estimation of the panel data fixed effect logit model (conditional logit approach for the parameter) and their corresponding average partial effects.

Semi-parametric Identification

- **Step 1: A fixed effects logit approach to identify and estimate the parameter β_0.** Use conditional logit transformation to eliminate the individual effect C by considering the joint distribution of $Y = (Y_1, .., Y_T)'$ conditional on X, C and $n = \sum_{t=1}^{T} Y_t$.

$$
\begin{aligned}
P(Y_1, ..., Y_T | X, C, n) &= \frac{P(Y_1, ..., Y_T | X, C)}{P(n|X, C)} \\
&= \frac{P(Y_1|X, C) \cdots P(Y_T|X, C)}{P(n|X, C)} \\
&= P(Y_1, ..., Y_T | X, n)
\end{aligned}
$$

This conditional distribution does not depend on C. Therefore, we can apply a standard CMLE method to identify and estimate β_0 without requiring any assumptions on C.

Fixed Effect Logit Estimation

- Consider the $T = 2$ case, where $n_i = y_{i1} + y_{i2}$ takes a value in $\{0, 1, 2\}$.
- For $n_i = 1$,

$$
\begin{aligned}
& P(y_{i2} = 1 | x_i, c_i, n_i = 1) \\
&= \frac{P(y_{i2} = 1, n_i = 1 | x_i, c_i)}{P(n_i = 1 | x_i, c_i)} \\
&= \frac{P(y_{i2} = 1, y_{i1} = 0 | x_i, c_i)}{P(y_{i1} = 0, y_{i2} = 1 | x_i, c_i) + P(y_{i1} = 1, y_{i2} = 0 | x_i, c_i)} \\
&= \frac{\Lambda(x_{i2}\beta + c_i)[1 - \Lambda(x_{i1} + c_i)]}{[1 - \Lambda(x_{i1}\beta + c_i)]\Lambda(x_{i2}\beta + c_i) + \Lambda(x_{i1}\beta + c_i)[1 - \Lambda(x_{i2}\beta + c_i)]} \\
&= \Lambda[(x_{i2} - x_{i1})\beta]
\end{aligned}
$$

This conditional distribution does not depend on c_i.

Semi-parametric Identification

- **Step 2: A plug-in nonparametric identification of $f_{C|\overline{X}}$.** Applying the law of the total probability yields

$$
\begin{aligned}
P_{Y|X}(y|x) &= \int_{\mathcal{C}} P_{Y|X}(y|x,c) f_{C|X}(c|x) dc \\
&= \int_{\mathcal{C}} \left(\prod_{t=1}^{T} P_{Y_t|X_t,C}(y_t|x_t,c) \right) f_{C|\overline{X}}(c|\overline{x}) dc \\
&= \int_{\mathcal{C}} \prod_{t=1}^{T} H\left(y_t, x_t\beta_0 + c\right) f_{C|\overline{X}}(c|\overline{x}) dc,
\end{aligned}
$$

where the parameter β_0 is identified in **Step 1**, $f_{C|X} = f_{C|\overline{X}}$ and $H\left(y_t, x_t\beta_0 + c\right) = \Lambda(x_t\beta_0 + c)^{y_t}[1 - \Lambda(x_t\beta_0 + c)]^{1-y_t}$.

Convolution Type Function

- This is convolution type function and the Fourier transform of the convolution of two functions is the product of their individual Fourier transforms.
- The convolution of f and g:

$$f * g(w) = \int f(w-c)g(c)dc.$$

- The Fourier transform of the convolution of two functions

$$\widehat{(f * g)} = \widehat{f} \cdot \widehat{g}.$$

Apply the Fourier transform

- The simple case,

$$\underbrace{\int_{\mathcal{X}_t} e^{i\xi x_t} P_{Y|X}(y|x)dx_t}_{\substack{\text{Fourier transform of } P_{Y|X};\\ \text{identified from sample}}}$$

$$= \underbrace{\int_{\widetilde{\mathcal{X}}_t} e^{i\xi \frac{\widetilde{x}_t}{\beta_0}} \prod_{t=1}^{T} H\left(y_t, \widetilde{x}_t\right) \frac{d\widetilde{x}_t}{\beta_0}}_{\substack{\text{Fourier transform of } \prod_{t=1}^{T} H\left(y_t, \widetilde{x}_t\right);\\ \text{identified from \textbf{Step 1}}}} \times \underbrace{\int_{\mathcal{C}} e^{-i\xi \frac{c}{\beta_0}} f_{C|\overline{X}}(c|\overline{x})dc}_{\widehat{f_{C|\overline{X}}} \equiv \text{Fourier transform of } f_{C|\overline{X}}} .$$

- The above equation implies that the Fourier transform of $f_{C|\overline{X}}$ is identified.

Apply the inverse Fourier transform

- **Assumption.** (Well Defined Fourier Transforms) Assume $\int_{\mathcal{X}_t} P_{Y|X}(y|x)dx_t < c_1 < \infty$ and $0 < \int_{\widetilde{\mathcal{X}_t}} \prod_{t=1}^{T} H(y_t, \widetilde{x}_t)\, d\widetilde{x}_t < c_2 < \infty$.

- Proposition (Fourier Inversion Formula)

Suppose f, $\widehat{f}$ are both integrable and f is continuous on $\mathbb{R}^n$. Then

$$f(x) = \frac{1}{(2\pi)^n} \int_{\mathbb{R}^n} e^{ix\cdot\xi} \widehat{f}(\xi) d\xi.$$

- Apply the inverse Fourier transform to the Fourier transform $\widehat{f_{C|\overline{X}}}(\xi)$ to obtain

$$f_{C|\overline{X}}(c|\overline{x}) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-i\xi c} \widehat{f_{C|\overline{X}}}(\xi) d\xi.$$

Average Partial Effects

- The marginal distribution of C is also identified by

$$f_C(c) = \int_{\overline{\mathcal{X}}} f_{C|\overline{X}}(c|\overline{x}) f_{\overline{X}}(\overline{x}) d\overline{x}.$$

- The ASF and APE are identified by averaging the response probability and the partial effect across the distribution of C respectively:

$$\mathsf{ASF}(x_t) = \int_{\mathcal{C}} \Lambda(x_t\beta_0 + c) f_C(c) dc,$$

$$\mathsf{APE}(x_{tk}) = \beta_{0k} \int_{\mathcal{C}} \frac{\partial \Lambda(x_t\beta_0 + c)}{\partial x_{tk}} f_C(c) dc.$$

Semiparametric Two-step Estimators

- The proposed estimation method closely follows the identification steps.
- **Step 1: A fixed effects logit estimate for β_0, $\widehat{\beta}_{fe}$.**
- **Step 2: A plug-in series estimator for APE.**
  Consider a series approximation for $f_{C|\overline{X}}(c|\overline{x})$ as follows

$$f^L_{C|\overline{X}}(c|\overline{x}) = \sum_{j=1}^{L}\sum_{k=1}^{L} a_{jk}\psi_j(c)\psi_k(\overline{x}),$$

where $\psi_j(c)$, and $\psi_k(\overline{x})$ are orthonormal series. Plugging $\widehat{\beta}_{fe}$ and the sieve expression into the identi. eq. in **Step 2** yields

$$\widehat{P}_{Y|X}(y|x) = \int_{\mathcal{C}} \prod_{t=1}^{T} H\Big(y_t, x_t\widehat{\beta}_{fe}+c\Big) f^L_{C|\overline{X}}(c|\overline{x})dc + o_p(1)$$

$$= \sum_{j,k=1}^{L} a_{jk} \int_{\mathcal{C}} \prod_{t=1}^{T} H\Big(y_t, x_t\widehat{\beta}_{fe}+c\Big)\psi_j(c)dc\cdot\psi_k(\overline{x}) + o_p(1).$$

A Plug-in Series Estimator for APE

- The density restriction, $\int f^L_{C|\overline{X}}(c|\overline{x})dc = 1$ for any $\overline{x}$, gives that $a_{11} = 1$ and $a_{1k} = 0$ for $k = 2, ..., L$.
- Plugging the density restrictions ($a_{11} = 1$ and $a_{1k} = 0$ for $k = 2, ..., L$) back and rearranging the term yields

$$\widehat{P}_{Y|X}(y|x) - H_1\Big(y, x; \widehat{\beta}_{fe}\Big)\psi_1(\overline{x}) = \sum_{j=2}^{L}\sum_{k=1}^{L} a_{jk} H_j\left(y, x; \widehat{\beta}_{fe}\right)\psi_k(\overline{x}) + o_p(1),$$

where $H_j\left(y, x; \widehat{\beta}_{fe}\right) \equiv \int_{\mathcal{C}} \prod_{t=1}^{T} H\left(y_t, x_t\widehat{\beta}_{fe} + c\right)\psi_j(c)dc$.

- The above equation can be written in a matrix notation and the coefficients a_{jk} have a closed form solution from a sample.

A Plug-in Series Estimator for APE

Suppose that there exists a panel data sample of size N, $\{y_{it}, x_{it}, \overline{x}_i\}$ for $i = 1, 2, ..., N; t = 1, 2, ..., T$. Define the following matrices

$$F_1 \equiv \begin{bmatrix} \widehat{P}_{Y|X}(y_1|x_1) - H_1\left(y_1, x_1; \widehat{\beta}_{fe}\right)\psi_1(\overline{x}_1) \\ \vdots \\ \widehat{P}_{Y|X}(y_N|x_N) - H_1\left(y_N, x_N; \widehat{\beta}_{fe}\right)\psi_1(\overline{x}_N) \end{bmatrix}_{N\times 1},$$

$$F_2 \equiv \begin{bmatrix} H_2\left(y_1, x_1; \widehat{\beta}_{fe}\right)\psi_1(\overline{x}_1) & \cdots & H_L\left(y_1, x_1; \widehat{\beta}_{fe}\right)\psi_L(\overline{x}_1) \\ \vdots & \cdots & \vdots \\ H_2\left(y_N, x_N; \widehat{\beta}_{fe}\right)\psi_1(\overline{x}_N) & \cdots & H_L\left(y_N, x_N; \widehat{\beta}_{fe}\right)\psi_L(\overline{x}_N) \end{bmatrix}_{N\times(L^2-L)}.$$

A Plug-in Series Estimator for APE

and

$$A_{L^2} \equiv \begin{bmatrix} a_{21} \\ \vdots \\ a_{2L} \\ \vdots \\ a_{jk} \\ \vdots \\ a_{LL} \end{bmatrix}_{(L^2-L)\times 1}.$$

With these matrices, we have

$$F_1 = F_2 A_{L^2} + o_p(1).$$

If the matrix F_2 is of full column rank, then the solution to the coefficient matrix is given by

$$\widehat{A_{L^2}}(\widehat{\beta}_{fe}) \equiv \left(F_2^T F_2\right)^{-1} F_2^T F_1.$$

A Plug-in Series Estimator for APE

- We have an estimated series estimator for $f_{C|\overline{X}}(c|\overline{x})$:

$$\widehat{f}^{L}_{C|\overline{X}}(c|\overline{x}) = \sum_{j=1}^{L}\sum_{k=1}^{L}\widehat{a_{jk}}\psi_j(c)\psi_k(\overline{x}).$$

- The marginal distribution f_C can be expressed in terms of

$$f_C(c) = \int f_{C|\overline{X}}(c|\overline{x})f_{\overline{X}}(\overline{x})d\overline{x} = \mathsf{E}_{\overline{X}}[f_{C|\overline{X}}(c|\overline{x})].$$

It follows that

$$\widehat{f}^{L}_{C}(c) = \frac{1}{N}\sum_{i=1}^{N}\widehat{f}^{L}_{C|\overline{X}}(c|\overline{x}_i) = \frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{L}\sum_{k=1}^{L}\widehat{a_{jk}}\psi_j(c)\psi_k(\overline{x}_i).$$

A Plug-in Series Estimator for APE

Plugging the fixed effects logit estimate $\widehat{\beta}_{fe}$ and the series estimator $\widehat{f}_C^L(c)$ back to the definition of ASF yields

$$\widehat{\mathsf{ASF}}(x_t)=\int_{\mathcal{C}}\Lambda(x_t\widehat{\beta}_{fe}+c)\widehat{f}_C^L(c)dc=\sum_{j,k=1}^{L}\widehat{a_{jk}}\left(\int_{\mathcal{C}}\Lambda\left(x_t\widehat{\beta}_{fe}+c\right)\psi_j(c)dc\right)\cdot\overline{\psi_k}.$$

Thus, the proposed plug-in series estimator for ASF has the following matrix expression

$$\begin{bmatrix}\int_{\mathcal{C}}\Lambda\left(x_t\widehat{\beta}_{fe}+c\right)\psi_1(c)dc\\ \vdots\\ \int_{\mathcal{C}}\Lambda\left(x_t\widehat{\beta}_{fe}+c\right)\psi_L(c)dc\end{bmatrix}^T_{L\times 1}\begin{bmatrix}\widehat{a_{11}} & \cdots & \widehat{a_{1L}}\\ \vdots & \cdots & \\ \widehat{a_{L1}} & \cdots & \widehat{a_{LL}}\end{bmatrix}_{L\times L}\begin{bmatrix}\overline{\psi_1}\\ \vdots\\ \overline{\psi_L}\end{bmatrix}_{L\times 1}.$$

Asymptotic Properties of the Two-Step Estimator

- (Consistency of $\widehat{\beta}_{fe}$) Let $\{(y_{it}, x_{it}) : i = 1, \ldots, N, t = 1, \ldots, T\}$ be an i.i.d. random sample across i. Let Θ be the parameter set and denote the parametric model of the conditional density as $\{P(y|x, \sum_{t=1}^{T} y_t; \beta) : \beta \in \Theta\}$. Suppose that (a) $P(y|x, \sum_{t=1}^{T} y_t; \beta)$ is a density with respect to y for all x and β ; (b) for some $\beta_0 \in \Theta$, $P(y|x, \sum_{t=1}^{T} y_t) = P(y|x, \sum_{t=1}^{T} y_t; \beta_0)$ for all x, and β_0 is the unique solution to $\max_{\beta \in \Theta} \mathsf{E}[l(w_i, \beta)]$; (c) Θ is a compact set; (d) for each β, $P(y|x, \sum_{t=1}^{T} y_t; \beta)$ is a Borel measurable function; (e) $l(w, \beta)$ is a continuous function on Θ for each w; and (f) $|l(w, \beta)| \leq b(w)$, for all β, and $\mathsf{E}[b(w)] < \infty$. Then $plim \widehat{\beta}_{fe} = \beta_0$.

Asymptotic Properties of the Two-Step Estimator

- (Asymptotic Normality of $\widehat{\beta}_{fe}$) Let the conditions of the consistency result of $\widehat{\beta}_{fe}$ hold. In addition, assume that (a) $\beta_0 \in$ int(Θ); (b) for each w, $l(w,\beta)$ is twice continuously differentiable on some neighborhood $\Theta_0 \in$ int(Θ); (c) the conditional support of y does not depend on the parameters β; (d) $\int \sup_{\beta\in\Theta_0} ||\nabla_\beta \exp l_i|| dw_i < \infty$ and $\int \sup_{\beta\in\Theta_0} ||\nabla^2_\beta \exp l_i|| dw_i < \infty$; and (e) $\mathsf{E}[s_i(\beta_0)s_i(\beta_0)']$ is positive definite. Then

$$\sqrt{N}(\widehat{\beta}_{fe} - \beta_0) = \left(N^{-1}\sum_{i=1}^{N} H_i(\breve{\beta})\right)^{-1}\left(-N^{-1/2}\sum_{i=1}^{N} s_i(\beta_0)\right)$$
$$\xrightarrow{d} N(0, \mathsf{E}[s_i(\beta_0)s_i(\beta_0)']^{-1}),$$

where $\breve{\beta}$ is on the line segment between $\widehat{\beta}_{fe}$ and β_0.

Asymptotic Properties of the Two-Step Estimator

- **Assumption A.** (a) With $\gamma > 1$, we have $f_1(\cdot|\cdot) \in \mathcal{A}$ under the norm $\|\alpha\|_{\Lambda^\gamma}$; (b) for any $\alpha \in \mathcal{A}$, there exists $\Pi_N \alpha \in \mathcal{A}_N$ such that $\|\Pi_N \alpha - \alpha\|_{\Lambda^\gamma} = o(1)$; (c) $L \to +\infty$ and $L/N \to 0$, where L is the order of the sieve approximation; and (d) both $\mathcal{C}$ and $\overline{\mathcal{X}}$ are compact.
- **(Consistency of $\widehat{f}^L_{C|\overline{X}}$)** Suppose that all assumptions in Asymptotic Normality of $\widehat{\beta}_{fe}$ and Assumption A hold. Let $\widehat{f}^L_{C|\overline{X}}(c|\overline{x})$ be the sieve estimator for $f_{C|\overline{X}}(c|\overline{x})$ in Step 2, then we have $\|\widehat{f}^L_{C|\overline{X}} - f_{C|\overline{X}}\|_{\Lambda^\gamma} = o_p(1)$.
- **(Consistency of $\widehat{\text{ASF}}(x_t^*)$)** Suppose that all assumptions in Asymptotic Normality of $\widehat{\beta}_{fe}$ and Assumption A hold. Then, the estimator $\widehat{f}^L_C$ satisfies $\|\widehat{f}^L_C - f_C\|_{\Lambda^\gamma} = o_p(1)$. In addition, the estimated $\widehat{\text{ASF}}(x_t^*)$ and $\widehat{\text{APE}}_j(x_t^*)$ converge to $\text{ASF}(x_t^*)$ and $\widehat{\text{APE}}_j(x_t^*)$ in probability, respectively.

Monte Carlo Study

- DGPs:

$$Y_t = 1\left(0.5X_{1t} + 0.5X_{2t} + C + E_t > 0\right), \quad \text{for } t = 1, 2, 3,$$

$$X_{1t} = 0.5X_{1t-1} + \varepsilon_{10}, X_{11} \sim U(0,1), \varepsilon_0 \sim \text{Trun}(N(0,1), [-5,5]),$$

$$X_{2t} \sim U(0,1), C = 0.5\overline{X}_1 + V, \ \ \overline{X}_1 = \frac{1}{3}\sum_{t=1}^{3} X_{1t},$$

E_t has a standard logistic distribution independent of X_t and C

- Consider four different types of DGPs for V as follows:

DGP I: $V = h(V)e, \ e \sim U(0,1)$ with $h(V) = 1$,

DGP II: $V = h(V)e, \ e \sim \text{Trun}(N(0,1), [0,1])$ with $h(V) = 1$,

DGP III: $V = h(V)e, \ e \sim U(0,1)$ with $h(V) = |\overline{X}_1|$,

DGP IV: $V = h(V)e, \ e \sim \text{Trun}(N(0,1), [0,1])$ with $h(V) = |\overline{X}_1|$.

Simulation Results of $N = 500$, $X = X_{1t}$

	$L=2$		$L=3$		$L=4$	
	$\widehat{\beta}_{fe}$	$\widehat{\text{ASF}}(x_t^*)$	$\widehat{\beta}_{fe}$	$\widehat{\text{ASF}}(x_t^*)$	$\widehat{\beta}_{fe}$	$\widehat{\text{ASF}}(x_t^*)$
DGP I:						
True	0.500	0.700	0.500	0.700	0.500	0.700
Mean	0.509	0.684	0.509	0.682	0.509	2.354
Median	0.511	0.684	0.511	0.682	0.511	2.297
RMSE	0.093	0.018	0.093	0.020	0.093	2.229
DGP II:						
True	0.500	0.692	0.500	0.692	0.500	0.692
Mean	0.503	0.683	0.503	0.681	0.503	2.374
Median	0.497	0.684	0.497	0.682	0.497	2.284
RMSE	0.103	0.010	0.103	0.012	0.103	2.290
DGP III:						
True	0.500	0.649	0.500	0.649	0.500	0.649
Mean	0.502	0.684	0.502	0.682	0.502	2.510
Median	0.507	0.683	0.507	0.683	0.507	2.448
RMSE	0.089	0.040	0.089	0.039	0.089	2.484
DGP IV:						
True	0.500	0.645	0.500	0.645	0.500	0.645
Mean	0.502	0.683	0.502	0.682	0.502	2.490
Median	0.497	0.684	0.497	0.682	0.497	2.410
RMSE	0.099	0.046	0.099	0.046	0.099	2.482

Simulation Results $N = 500$, $X = [X_{1t}, X_{2t}]$

	$L=2$			$L=3$			$L=4$		
	$\widehat{\beta}_{fe1}$	$\widehat{\beta}_{fe2}$	$\widehat{\text{ASF}}(x_t^*)$	$\widehat{\beta}_{fe1}$	$\widehat{\beta}_{fe2}$	$\widehat{\text{ASF}}(x_t^*)$	$\widehat{\beta}_{fe1}$	$\widehat{\beta}_{fe2}$	$\widehat{\text{ASF}}(x_t^*)$
DGP I:									
True	0.500	0.500	0.692	0.500	0.500	0.692	0.500	0.500	0.692
Mean	0.509	0.455	0.699	0.517	0.456	0.687	0.516	0.463	2.672
Median	0.506	0.446	0.682	0.518	0.457	0.685	0.521	0.461	2.038
RMSE	0.098	0.255	0.040	0.099	0.270	0.026	0.098	0.256	3.231
DGP II:									
True	0.500	0.500	0.684	0.500	0.500	0.684	0.500	0.500	0.684
Mean	0.505	0.491	0.699	0.510	0.493	0.680	0.502	0.489	2.967
Median	0.515	0.468	0.684	0.516	0.494	0.682	0.512	0.487	2.197
RMSE	0.085	0.265	0.043	0.087	0.252	0.035	0.088	0.259	4.035
DGP III:									
True	0.500	0.500	0.641	0.500	0.500	0.641	0.500	0.500	0.641
Mean	0.512	0.479	0.691	0.512	0.460	0.672	0.513	0.462	2.880
Median	0.504	0.481	0.679	0.510	0.445	0.671	0.507	0.459	2.010
RMSE	0.088	0.277	0.058	0.097	0.254	0.038	0.096	0.251	3.527
DGP IV:									
True	0.500	0.500	0.637	0.500	0.500	0.636	0.500	0.500	0.636
Mean	0.518	0.481	0.691	0.512	0.504	0.669	0.516	0.489	2.888
Median	0.528	0.488	0.677	0.511	0.496	0.670	0.510	0.477	1.879
RMSE	0.091	0.248	0.058	0.091	0.242	0.039	0.089	0.236	4.060

The Effects of Income on the Labour Force Participation of Older Women

- Use data from the Medical Expenditure Panel Survey (MEPS) Panel 20 in 2015 and 2016. The sample consists of two periods of 2,837 women aged older than 50.
- The latent utility for the labor force participation decision is derived from the following:

$$\text{LFP}^*_{it} = X_{it}\beta_0 + C_i + E_{it} \quad i = 1, \ldots, N; t = 1, \ldots, T$$

where C_i represents as unobserved individual skill level or incentive for work and E_{it} is an logistic idiosyncratic error term. An individual decide to work, $\text{LFP}_{it} = 1$, if

$$X_{it}\beta_0 + C_i + E_{it} > 0 \text{ or } X_{it}\beta_0 + E_{it} > -C_i.$$

Sample Statistics of MEPS

Variable	Definition		Mean	Std. Dev.
LFP	=1 if the person is employed;	Overall	0.411	0.492
	0 otherwise	Between		0.477
		Within		0.122
Age	Age older than 50	Overall	64.406	10.122
		Between		10.111
		Within		0.487
Income	Family income	Overall	60,100	59,733
		Between		56,294
		Within		19,989
Good health	=1 if self-rated health is excellent	Overall	0.454	0.498
	or very good;0 otherwise	Between		0.435
		Within		0.242
Marriage	=1 if the person is married;	Overall	0.473	0.499
	0 otherwise	Between		0.496
		Within		0.060
Sample size	N: cross-sectional unit		2,837	
	T: time unit		2	

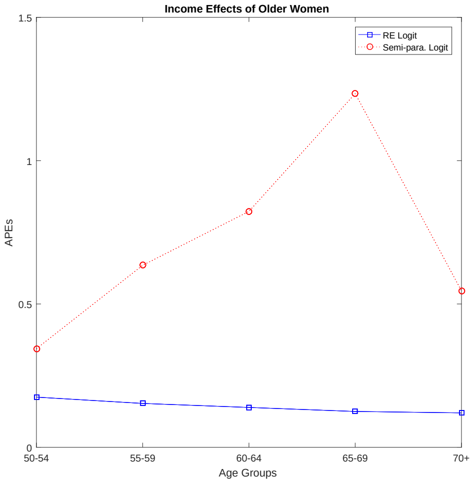
The Effects of Income on the Labour Force Participation

- Four empirical questions:
  1. Whether the relationship between unobserved time invariant factors and income affects on the effect of income on the labour force participation of older people
  2. Whether there are different effects of income across age groups
  3. Whether the potential bias due to the unobserved heterogeneity increases for older age groups
  4. Whether health status and martial status have impacts on labor force behavior after controlling for the unobserved heterogeneity.
- The modelling of the unobserved individual fixed effect:
  1. Pooled RE Logit: Restriction $D(C|X)$ independence
  2. Pooled CRE Logit: Restriction $D(C|X)$ linear mean; $C = \lambda\overline{X} + V$
  3. Semi-parametric Logit: Restriction $f_{C|X} = f_{C|\overline{X}}$

Panel Logit Estimates of Labor Force Participation

	RE Logit		CRE Logit		Semi-para. Logit	
	Coefficient	APE	Coefficient	APE	Coefficient	APE
Age 55-59	1.259	0.185	4.959**	0.540**	-7.804	-3.317***
	(1.635)	(0.240)	(2.234)	(0.244)	(6.401)	(0.188)
Age 60-64	1.625	0.239	7.052***	0.769***	-28.664	-12.182***
	(1.667)	(0.245)	(2.238)	(0.244)	(1667.386)	(0.705)
Age 65-69	1.705	0.251	7.78***	0.848***	-39.285	-16.698***
	(1.796)	(0.264)	(2.327)	(0.254)	(1667.408)	(0.968)
Age 70 and older	0.469	0.069	9.933***	1.083***	-21.93	-9.319***
	(2.016)	(0.296)	(2.57)	(0.280)	(1667.386)	(0.601)
Income age 50-54	1.192***	0.175***	0.918***	0.100***	0.808**	0.344***
	(0.128)	(0.019)	(0.193)	(0.021)	(0.384)	(0.019)
Income age 55-59	1.037***	0.153***	0.461***	0.050***	1.494***	0.635***
	(0.102)	(0.015)	(0.126)	(0.014)	(0.588)	(0.033)
Income age 60-64	0.944***	0.139***	0.281**	0.031**	1.937**	0.823***
	(0.102)	(0.015)	(0.12)	(0.013)	(0.955)	(0.043)
Income age 65-69	0.847***	0.125***	0.191	0.021	2.902**	1.234***
	(0.117)	(0.017)	(0.13)	(0.014)	(1.226)	(0.064)
Income age 70+	0.813***	0.120***	-0.01	-0.001	1.285	0.546***
	(0.144)	(0.021)	(0.157)	(0.017)	(0.934)	(0.029)
Good health	0.759***	0.112***	0.043	0.005	0.113	0.048***
	(0.083)	(0.012)	(0.09)	(0.010)	(0.35)	(0.003)
Marriage	-0.806***	-0.119***	-0.734**	-0.080**	-18.613	-9.518***
	(0.106)	(0.0156)	(0.349)	(0.038)	(2765.882)	(0.486)

The Estimated APEs of Income of Older Women



Conclusion

- This paper develops semiparametric estimation methods for the panel data logit models and their corresponding average partial effects without fully specifying the distribution of the individual effect C.
- This paper addresses two issues of panel data logit models. (1) new identification conditions for APEs of panel data logit models, and (2) a data-driven specification of conditional distributions of the unobserved heterogeneity.
- The advantage of the proposed approach: (1) internally consistent modelling to integrate out the FE logit estimate and the unobserved heterogeneity; (2)all identification assumptions are related to the regularity conditions of the proposed nonlinear panel data models and observable density functions. No high level assumptions; and (3) the method can be extended to dynamic panel data models and other regression panel data models.

Conclusion

- Our empirical application on labor force participation of older workers shows that the unobserved individual heterogeneity plays an important role in older workers' employment status, and the proposed semiparametric estimator can substantially reduce the potential biases arising from parametric assumptions.
- The estimation result provides evidence that the influence of unobserved heterogeneity on both income and labor force participation increases over time and the income gradient for older age groups is more sensitive to bias due to the unobserved heterogeneity.
- The proposed approach may provide a solution to estimate APEs to other nonlinear panel data models with a linear latent variable structure such as $X_t\beta_0 + C$.