Estimating Semi-Parametric Panel Multinomial Choice Models Using Cyclic Monotonicity[1]

Xiaoxia Shi [1] Matthew Shum [2] Wei Song [3]

[1]Wisconsin [2]CalTech [3]Xiamen University

May 29, 2020

[1]Presentation by Manu Navjeevan

Introduction

Paper proposes a new semi-parametric identification and estimation approach to multinomial choice models in a panel data setting.

- General approach is based on cyclic monotonicity
- This helps derive identifying inequalities without requiring shape restrictions for the distribution of the underlying shock

Intuition underlying the paper: Agents who change their decisions over time must be better of by doing so. If the underlying utility structure is constant over time, this can give us identifying information on the aforementioned structure.

Definition 1 (Cyclic Monotonicity)

Consider a function $f : \mathbb{U} \rightarrow \mathbb{R}^k$ where $\mathbb{U} \subseteq \mathbb{R}^k$ and a lenth m cycle of points in $\mathbb{R}^k : u_1, u_2, \ldots, u_m, u_1$. The function f is cyclic monotone if

$$\sum_{m=1}^{M} (u_m - u_{m+1})' f(u_m) \geq 0 \tag{1}$$

for all length m cycles on its domain.

Cyclic monotonicity and the following proposition will be exploited to generate moment inequality restrictions on the data.

Proposition 1

Consider a differentiable function $F : \mathbb{R} \to \mathbb{R}$ for an open, convex, set $\mathbb{U} \subseteq \mathbb{R}^K$. If F is convex on $\mathbb{U}$ then the gradient $\nabla F(u) = \partial F(u)/\partial u$ is cyclic monotone on $\mathbb{U}$

Proof of the proposition should follow quickly from the fact that $\nabla^2 F(u)$ is positive semidefinite. This proposition will be used later on to generate restrictions on the model parameters.

Paper considers a panel multinomial choice problem. Agent i chooses from $K+1$ options $k=0,\ldots,K$. Choosing option k in period t gives the agent indirect utility

$$\underbrace{\beta' X_{it}^k}_{d_x-\text{dimensional covariates}} + \underbrace{A_i^k}_{\text{fixed effect}} + \underbrace{\epsilon_{it}^k}_{\text{unobserved shock}} \tag{2}$$

Agent chooses option that gives them the highest utility

$$Y_{it} = \mathbb{1}\{\beta' X_{it}^k + A_i^k + \epsilon_{it}^k \geq \beta' X_{it}^{k'} + A_i^{k'} + \epsilon_{it}^{k'}, \forall k'\} \tag{3}$$

Let U_{it}^{k} denote the systemic component of the utility, i.e

$$U^k = \beta' X^k + A_i^k$$

Let $\mathbf{U} = (U^1, \ldots, U^K)$ be the random vector of systemic utilities and let $\mathbf{u}$ denote a specific realization of $\mathbf{U}$.

$$W(\mathbf{u}) = \mathbb{E}\left[\max_{k=0,1,\ldots,K}\left[U^k + \epsilon^k\right]\middle|\mathbf{U} = \mathbf{u}\right] \tag{4}$$

Note that via convexity of maximum and linearity of indirect utility, $W(\cdot)$ is convex. This motivates the following Lemma and application of cyclic monotnoicity.

Multinomial Choice Model

Applying Cyclic Monotonicity

Lemma 1

Suppose that $\mathbf{U}$ is independent of ϵ and that the distribution of ϵ is absolutely continuous with respect to the Lebesgue measure. Then:

1. $W(\cdot)$ is convex on $\mathbb{R}^K$
2. $W(\cdot)$ is differentiable on $\mathbb{R}^K$,
3. $\mathbf{p}(\mathbf{u}) = \nabla W(\mathbf{u})$. where $\mathbf{p}(\mathbf{u}) = \mathbb{E}[Y|\mathbf{U} = \mathbf{u}]$[2].
4. $\mathbf{p}(\mathbf{u})$ is cyclic monotone on $\mathbb{R}^K$

This lemma allows us to apply restrictions generated by cyclic monotonicity to the conditional choice probabilities to develop restrictions on model parameters.

[2] This is the observed conditional choice probability

Two Period Panel Model

Assumptions for Partial Identification

Consider a two period model as above, where indirect utility for choice k in each time period is given

$$\beta' X_{it}^k + A_i^k + \epsilon_{it}^k$$

for $t = 1, 2, k = 0, 1, \ldots, K$. As before, let Y_{it}^k denote the indicator for choice k

Assumption 1

Assume that

1. The random vectors ϵ_{i1} and ϵ_{i2} are identically distributed conditional on A_i, X_{i1}, X_{i2}
2. The conditional distribution of ϵ_{it} given A_i, X_{i1}, X_{i2} is absolutely continuous with respepect to the Lesbegue measure for $t = 1, 2$ everywhere on the support of A_i, X_{i1}, X_{i2}

Previous lemma established cyclical monotonicity of the contitional choice probability, $\mathbb{E}[Y^k|X_1, X_2]$.

$$\left(\mathbb{E}\left[Y_{i1}'|X_{i1}, X_{i2}\right] - \mathbb{E}\left[Y_{i2}'|X_{i1}, X_{i2}\right]\right)(X_{i1}'\beta - X_{i2}'\beta) \geq 0 \qquad (5)$$

or, rewritten

$$\mathbb{E}[\Delta Y_i'|X_{i1}, X_{i2}]\Delta X_i'\beta \geq 0 \qquad (6)$$

Steps

This set of inequality constraints can be used to place useful moment restrictions on β. In order to achieve point identification, need the right sort of variation in these moment inequalities so that they form moment *equalities*.

Two Period Panel Model

Point Identification Regularity Conditions

For point identification, want restructions that ensure that $\text{supp}(\Delta X_i \mathbb{E}[\Delta Y_i | X_{i1}, X_{i2}])$ is "rich enough"

Assumption 2

1. The conditional support of $\epsilon_{it} | A_i, X_{i1}, X_{i2}$ is $\mathbb{R}^K$ with positive probability everywhere
2. The conditional distribution of $(\epsilon_{it} + A_i)$ given $(X_{i1}, X_{i2}) = (x_1, x_2)$ is uniformly continuous[3]

We'll also need a condition on X that ensures that X has enough variation to make suitable enough comparasions

---

[3] This is a suffecient condition for the continuity of the function $\mathbb{E}[\Delta Y_i | X_{i1}, X_{i2}]$.

Two Period Panel Model

Motivating Necessary Variation in X

By definition

$$\Delta X_i E[\Delta Y_i | X_{i1}, X_{i2}] = \sum_{k=1}^{K} \Delta X_i^k E[\Delta Y_i^k | X_{i1}, X_{i2}]$$

Want suffecient variation of this term. Hard to do because it is a weighted sum, weighted by a non-primitive. To this end define

$$\begin{aligned} G_I &= \bigcup_k \text{supp}(\Delta X_i^k | \Delta X_i^{-k} = 0) \\ G_{II} &= \bigcup_k \text{supp}(\Delta X_i^k | \Delta X_i^k = \Delta X_i^1, \forall k) \\ G &= G_I \cup G_{II} \end{aligned}$$

Identifying restrictions will then be placed on G.

Two Period Panel Model

Point Identifying Assumption

Either of the following Assumptions, when combined with the prior assumptions, are suffecient for identification of β

Assumption 3

The set G contains an open R^{d_x} ball around the origin.

Let r.v $g = (\Delta X_i \mathbb{E}[Y_i | X_{i1}, X_{i2}])$. Let g_{-j} denote g with the j-th element removed and $G_j(g_{-j}) = \{g_j \in \mathbb{R} : (g_j, g'_{-j})' \in G\}$

Assumption 4

For some $j^* \in \{1, 2, \ldots, d_x\}$:

1. $G_{j^*}(g_{-j^*}) = \mathbb{R}$ for all g_{-j^*} in a subset G^0_{-j} of G_{-j^*}
2. $G^0_{-j^*}$ is not contained in a proper linear subspace of $\mathbb{R}^{d_x - 1}$
3. the j^*-th of β, denoted β_{j^*} is nonzero.

Examples

Identification result is stated using the following criterion function:

$$Q(b) = \mathbb{E}\left|\min\left(0, \mathbb{E}[\Delta Y_i' | X_{i1}, X_{i2}]\Delta X_i' b\right)\right| \tag{7}$$

Which will be returned to in considering estimation.

Theorem 1

Under Assumptions 1,2, and either 3 or 4, we have $Q(\beta) = 0$ and $Q(b) > 0$ for all $b \neq \beta$ such that $b = \mathbb{R}^{d_x}$ and $\|b\| = 1$

Prior criterion function sets up GMM estimator of β, $\hat{\beta} = \bar{\beta}/\|\bar{\beta}\|$ where

$$\bar{\beta} = \arg \min_{b \in \mathbb{R}^{d_X} : \max_j |b_j| = 1} n^{-1} \sum_{i=1}^{n} \left[\left(b' \Delta X_i \right) \left(\Delta \hat{p}(X_{i1}, X_{i2}) \right) \right]_{-}$$

here $\Delta \hat{p}(X_{i1}, X_{i2}) = \hat{p}_2(X_{i1}, X_{i2}) - \hat{p}_1(X_{i1}, X_{i2})$ and $\hat{p}_t(x_1, x_2)$ is a uniformly consistent estimator for $\mathbb{E}(Y_{it}|X_{i1}, X_{i2})$

Assumption 5

Assume that:

1. $\max_i \|\hat{p}(\cdot) - p(\cdot)\| \to_p 0$ is uniformly consistent and
2. $\max_{t=1,2} \mathbb{E}[\|X_{it}\|] < \infty$

Theorem 2

(Consistency) Under Assumptions 1, 2, 5 and either 3 or 4:

$$\hat{\beta} \xrightarrow{p} \beta \text{ as } n \to \infty$$

Further Results

- Paper also considers what to do in longer panels
  - Cylic monotonicity becomes more complicated in these models, as more comparasions can be made between time periods.
  - Theoretically, this should relax identification conditions and make estimation more effecient.
  - Longer panel adds more moment restrictions and should make estimation more effecient at the least.
- Model can be applied to a model with aggregate data, following very similar asusmptions and logic.
- Cyclic monotonicity can also be applied to a cross sectional model without fixed effects. In this case the identification results with cyclic monotonicity reduce to those of Manski (1975) and Han (1987)

Discussion and Conclusion

- Paper provides identification of a semiparametric multinomial choice model without placing shape restrictions on the data
- Identification follows from assuming a constant underlying utility structure and then looking at how people change their decisions over time.
  - Formally shown through cyclic monotonicity
- Paper provides conditions for partial identification, point identification, and consistent parameter estimation

Begin by letting η be a K-dimensional vector with k-th element η^k and define

$$\mathbf{p}(\eta, x_1, x_2, a) := \left(\mathbb{P}\left[\epsilon_{i1}^{k} + \eta^{k} \geq \epsilon_{i1}^{k'} + \eta^{k'} \ \forall k' | X_{i1} = x_1, X_{i2} = x_2, A_i = a\right]\right)_{\forall k}$$

Assumption 1.1 implies that

$$\mathbf{p}(\eta, x_1, x_2, a) := \left(\mathbb{P}\left[\epsilon_{i2}^{k} + \eta^{k} \geq \epsilon_{i2}^{k'} + \eta^{k'} \ \forall k' | X_{i1} = x_1, X_{i2} = x_2, A_i = a\right]\right)_{\forall k}$$

Assumption 1.2 along with Lemma 1 imply that $p(\eta, x_1, x_2, a)$ is cyclic monotone in η for all possible values of x_1, x_2. Using cyclic monotonicity for length 2 cycles we obtain, for any η_1, η_2 and x_1, x_2, a we have

$$(\eta_1 - \eta_2)' \left[p(\eta_1, x_1, x_2, a) - p(\eta_2, x_1, x_2, a) \right] \geq 0$$

Now let $\eta_1 = X_{i1}'\beta + A_i$ and $\eta_2' = X_{i2}'\beta + A_i$. By the definition of $p(\cdot)$ we have

$$p\left(X_{it}'\beta + A_i, X_{i1}, X_{i2}, A_i\right) = \mathbb{E}\left[Y_{it} | X_{i1}, X_{i2}, A_i\right]$$

Combining the above we have that

$$\left(\mathbb{E}\left[Y'_{i1}|X_{i1}, X_{i2}, A_i\right] - \mathbb{E}\left[Y'_{i2}|X_{i1}, X_{i2}, A_i\right]\right)\left(X'_{i1}\beta - X'_{i2}\beta\right) \geq 0$$

Fixed effect within the second paranthetical term on LHS defferences out. So can take conditional expectation given the X values to obtain

$$\left(\mathbb{E}\left[Y'_{i1}|X_{i1}, X_{i2}\right] - \mathbb{E}\left[Y'_{i2}|X_{i1}, X_{i2}\right]\right)\left(X'_{i1}\beta - X'_{i2}\beta\right) \geq 0$$

Back

Example 1: $\text{supp}((X_{it}^k)_{t=1,2,k=1,2}) = [0,1]^8$ then $\text{supp}((\Delta_i^k)_{k=1,2}) = [-1,1]^4$. In this case Assumption 3 is clearly satisfied.

Example 2: Suppose that the first covariate is a time dummy: $X_{1,it}^k = t$ for all k, t and the second covariate has unbounded support $\text{supp}((X_{2,it}^k)_{t=1,2,k=1,2})) = (c, \infty)^4$ for some $c \in \mathbb{R}$. Then,

$$\text{supp}(\Delta X_i^1 | \Delta X_i^1 = \Delta X_i^2) = \{1\} \times \mathbb{R}$$

so $G \supseteq G_{II} = \{-1,1\} \times \mathbb{R}$. Let $j^* = 2$ and $G_{-2}^0 = \{-1,1\}$. Then Assumption 4.2 holds and Assumption 4.1 holds becouse $G_2(-1) = G_2(1) = \mathbb{R}$. Assumption 4.3 holds as long as $\beta_2 \neq 0$

Back