Point-Identification in Simple Dynamic Binary Outcome Models

Bo E. Honoré with various co-authors

April, 2019

“Fixed” Effects in Discrete Choice Models

- An on-off obsession.
- Past/ongoing/future work with Kyriazidou/de Paula/Toft Hansen.

Combination of Well Understood Topics:

- Simple Panel Data Econometrics.
- Simple Binary Outcome Econometrics.
- Simple Time Series Econometrics.

But we still don’t understand it.

Outline

- A bit of survey.
  - Very, very selective.
- Current work.
  - Some new results.
  - Some puzzles.
- Try to convince you that this literature can be useful even if you don't care about "fixed" effects models.

Background (Big Picture)

A dependent variable (or vector) in time period t, y_{it}, is modelled as

$$y_{it} \sim f\left(\cdot \mid x_i^t, y_i^{t-1}, \alpha_i; \theta\right)$$

where x_{it} is a vector of explanatory variables and α_i, captures unobserved heterogeneity.

Alternatively (better)

$$y_{it} \sim f\left(\cdot \mid x_i^T, y_i^{t-1}, \alpha_i; \theta\right).$$

- Interested in identification of θ from the distribution of y_i^T given x_i^T for a small (finite) number of time periods, T.
- Nonparametric in the relationship between the heterogeneity term α_i and x_i^T.
- This is sometimes referred to as a "fixed effects" approach.

Disclaimer

Knowing θ in

$$y_{it} \sim f\left(\cdot \,|\, x_i^t, y_i^{t-1}, \alpha_i; \theta\right)$$

is typically not sufficient for calculating counterfactual distributions.

These are typically not point-identified even if θ is.

In principle not too hard to bound “average effects.”

Simplest Case: Static Panel Logit Models

Rasch (1960) considered a static panel data version of the standard logit model,

$$P\left(y_{it}=1|x_i^T,y_i^{t-1},\alpha_i\right)=\frac{\exp\left(x_{it}'\beta+\alpha_i\right)}{1+\exp\left(x_{it}'\beta+\alpha_i\right)}=\Lambda\left(x_{it}'\beta+\alpha_i\right).$$

In this case, the distribution of $(y_{i1},\ldots,y_{iT})$ conditional on $(x_{i1},\ldots,x_{iT})$ and on $\sum_{t=1}^{T} y_{it}$ does not depend on α_i, but for $T\geq 2$ it does depend on β.

The corresponding conditional likelihood can be used to identify and estimate β.

Very (or not) specific to logit assumption.

Much More Interesting Than I Make It Sound

- You might say that the static logit model is (kind of) boring and old-fashioned.
  - Indeed the conditioning approach dates back to Rasch (1960).
- This setup is at the core of three of the four (much sexier) papers in the econometrics session yesterday.
  - Botosaru and Muris (2019) uses the insight heavily.
- Recent new paper by Aguirregabiria, Gu, and Luo (2019) use the same conditioning ideas in a structural dynamic model.

2nd Simplest Case: AR(1) Logit Models

$$P\left(y_{it} = 1 | y_i^{t-1}, \alpha_i\right) = \frac{\exp\left(y_{it-1}\gamma + \alpha_i\right)}{1 + \exp\left(y_{it-1}\gamma + \alpha_i\right)}. \tag{1}$$

Conditional on $\left(y_{i1}, \sum_{t=1}^{T} y_{it}, y_{iT}\right)$, the distribution of $(y_{i1}, \ldots, y_{iT})$ does not depend on α_i, but for $T \geq 4$, it does depend on γ.

The corresponding conditional likelihood can therefore be used to identify and estimate γ.

(Cox (1958) and Chamberlain (1985))

3rd Simplest Case: AR(2) Logit Models

$$P\left(y_{it}=1|y_i^{t-1},\alpha_i,\gamma_{i1}\right)=\frac{\exp\left(\gamma_{i1}y_{it-1}+\gamma_2 y_{it-2}+\alpha_i\right)}{1+\exp\left(\gamma_{i1}y_{it-1}+\gamma_2 y_{it-2}+\alpha_i\right)}. \tag{2}$$

When the coefficient on y_{it-2} is 0, this corresponds to a Markov switching model.

The sufficient statistic for (γ_{i1},α_i) with $T\geq 6$ is $(y_{i1},y_{i2},s_{i1},s_{i11},y_{iT-1},y_{iT})$, where $s_{i1}=\sum_{t=1}^{T}y_{it}$ and $s_{i11}=\sum_{t=2}^{T}y_{it}y_{it-1}$.

This can be used to estimate γ_2 (can depend on y_{it-1})

(Chamberlain (1985), Magnac (2000))

There Is a Lot That We Do Not Know

$$P\left(y_{it} = 1 | y_i^{t-1}, \alpha_i, \gamma_{i1}\right) = \frac{\exp\left(\gamma_{i1} y_{it-1} + \gamma_2 y_{it-2} + \alpha_i\right)}{1 + \exp\left(\gamma_{i1} y_{it-1} + \gamma_2 y_{it-2} + \alpha_i\right)}.$$

What we can learn about the coefficient on y_{it-1} if we restrict it to be constant across individuals: $\gamma_{i1} = \gamma_1$ for all i?

- As far as I know, this is not known to be identified for any finite T.

Clearly a Toy Model

- But one that we should understand!

Numerical Strategy

- Assume a data generating process.
- Calculate the "true" probability of every sequence.
- Fix the identified parameters at their actual values.
- For a range of the possibly unidentified parameters, search over distributions of α_i trying to match the "true" probabilities.
  - Apply discretization. Current work with de Paula improves on this.
- Keep the parameters for which we get a match.

Try a bunch of data generating processes.

Figure: Identified Region when $\gamma_1 = -1$ and $\gamma_2 = -1$

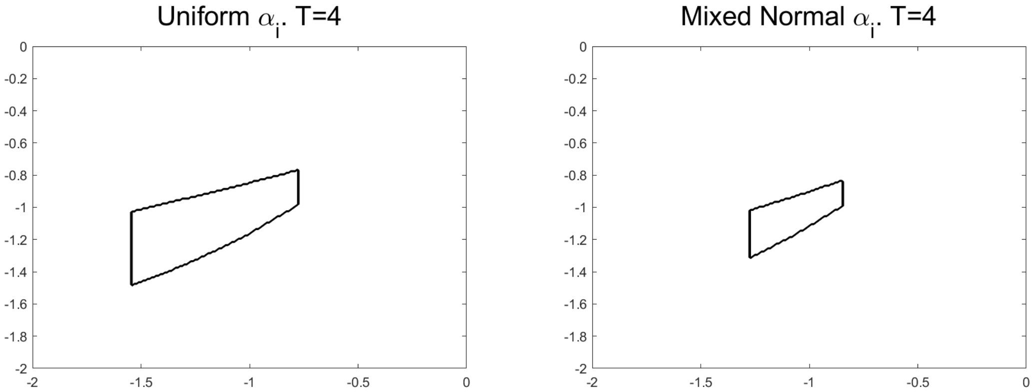


Table: Identified Regions in an AR(2) logit with $T = 5$. $\gamma_1 = \pm 1.5$, $\gamma_2 = \pm 0.7$

	Source of identification			
(γ_1, γ_2)	$(y_3, y_4, y_5 \mid 0, 0)$	$(y_3, y_4, y_5 \mid 1, 0)$	$(y_3, y_4, y_5 \mid 0, 1)$	$(y_3, y_4, y_5 \mid 1, 1)$
$\gamma_1 = -1.5$	$\{-1.5000\}$	$\{-1.5000\}$	$\{-1.5000\}$	$\{-1.5000\}$
$\gamma_2 = -0.7$	$[-0.7032, -0.6978]$	$\{-0.7000\}$	$\{-0.7000\}$	$[-0.7044, -0.6971]$
$\gamma_1 = -1.5$	$\{-1.5000\}$	$\{-1.5000\}$	$\{-1.5000\}$	$\{-1.5000\}$
$\gamma_2 = 0.7$	$[0.6985, 0.7010]$	$\{0.7000\}$	$\{0.7000\}$	$[0.6974, 0.7017]$
$\gamma_1 = 1.5$	$\{1.5000\}$	$\{1.5000\}$	$\{1.5000\}$	$\{1.5000\}$
$\gamma_2 = -0.7$	$[-0.7012, -0.6983]$	$\{-0.7000\}$	$\{-0.7000\}$	$[-0.7011, -0.6983]$
$\gamma_1 = 1.5$	$\{1.5000\}$	$\{1.5000\}$	$\{1.5000\}$	$\{1.5000\}$
$\gamma_2 = 0.7$	$[0.6981, 0.7029]$	$\{0.7000\}$	$\{0.7000\}$	$[0.6979, 0.7031]$

Identified?

If so, it seems that the proof is not by standard conditioning arguments.

- With $T = 5$, there are only eight sequences conditional on (y_1, y_2).

If not, it seems that the identified region is really, really small.

This is clearly a "toy" model. But....

Bivariate Models

Following Schmidt and Strauss (1975), consider the bivariate panel data model

$$P\left(y_{1,it} = 1 \middle| y_{2,it}, y_{1,i}^{t-1}, y_{2,i}^{t-1}, x_{1,i}^{T}, x_{2,i}^{T}, \alpha_{1,i}, \alpha_{2,i}\right) = \Lambda\left(\alpha_{1,i} + x_{1,it}'\beta_1 + \rho y_{2,it}\right)$$

and

$$P\left(y_{2,it} = 1 \middle| y_{1,it}, y_{1,i}^{t-1}, y_{2,i}^{t-1}, x_{1,i}^{T}, x_{2,i}^{T}, \alpha_{1,i}, \alpha_{2,i}\right) = \Lambda\left(\alpha_{2,i} + x_{2,it}'\beta_2 + \rho y_{1,it}\right).$$

Honoré and Kyriazidou (2018) show that in this case β_1, β_2 and ρ are identified with $T = 2$.

Entry Games

(with Aureo de Paula; will just give you the flavor of what we are doing)

In a given market in a given period

$$\begin{aligned} y_1 &= 1\left\{x_1'\beta - \gamma y_2 + \alpha_1 + \varepsilon_1 > 0\right\} \\ y_2 &= 1\left\{x_2'\beta - \gamma y_1 + \alpha_2 + \varepsilon_2 > 0\right\} \end{aligned}$$

$(\varepsilon_1, \varepsilon_2)$ independent logistic (actually, the distribution does not matter).

These could also be "social interactions." (Cameiro, Flores, Galasso, Ginja and de Paula (2019)).

With two time periods, and x market specific (so $x_{1jt} = x_{2jt} = x_{jt}$), we have

$$P\left(N^{j1}=2\right) = \frac{\exp\left(x_{j1}'\beta-\gamma+\alpha_{j1}\right)}{1+\exp\left(x_{j1}'\beta-\gamma+\alpha_{j1}\right)}\frac{\exp\left(x_{j1}'\beta-\gamma+\alpha_{j2}\right)}{1+\exp\left(x_{j1}'\beta-\gamma+\alpha_{j2}\right)}$$

$$P\left(N^{j2}=2\right) = \frac{\exp\left(x_{j2}'\beta-\gamma+\alpha_{j1}\right)}{1+\exp\left(x_{j2}'\beta-\gamma+\alpha_{j1}\right)}\frac{\exp\left(x_{j2}'\beta-\gamma+\alpha_{j2}\right)}{1+\exp\left(x_{j2}'\beta-\gamma+\alpha_{j2}\right)}$$

Clearly $P\left(N^{j1}=2\right) \gtreqqless P\left(N^{j2}=2\right)$ iff $x_{j1}'\beta \gtreqqless x_{j2}'\beta$.

Look at market where $N^{j1}=2$ or $N^{j2}=2$ but not both. Apply maximum score to the event $N^{j1}=2$ with $x_{j1}'-x_{j2}'$ as explanatory variable.

Logistic does not matter.

- x's not market specific.
- Conditioning on $x_{2j1} = x_{2j2} = x_{2j}$ (player 2 has the same x in two periods).

In that case,

$$P\left(N^{jt} = 2 \middle| \{x_{1js}, x_{2js}\}_{s=1}^{2}, \alpha_{1j}, \alpha_{2j}\right) = F\left(x'_{1jt}\beta - \gamma + \alpha_{j1}, x'_{2j}\beta - \gamma + \alpha_{j2}\right)$$

so

$$P\left(N^{j1} = 2 \middle| \{x_{1js}, x_{2js}\}_{s=1}^{2}, \alpha_{1j}, \alpha_{2j}\right) \gtreqless P\left(N^{j2} = 2 \middle| \{x_{1js}, x_{2js}\}_{s=1}^{2}, \alpha_{1j}, \alpha_{2j}\right)$$

$$\text{iff} \qquad x'_{1j1}\beta \gtreqless x'_{1j2}\beta.$$

As a result, we can identify and estimate β by conditioning on markets where $N^{j1} = 2$ or $N^{j2} = 2$, but not both.

What about γ?

- Have not been able to find conditioning arguments that are useful for point-identifying γ.
- Plan to investigate the numerical region for γ using the same general strategy as above.

Step forward in terms of calculation:

- Before we thought of the discretization as an approximation.
- We will exploit that there is no loss of generality in assuming that α is discrete with a small (known) number of unknown points of support (Theorem).

Hope to prove that values of the points of support can be considered known.

Dynamics

Narendranthan, Nickell, and Metcalf (1985):

$$
\begin{aligned}
y_{1,it} &= 1\left\{y_{1,it-1}\gamma_{11} + y_{2,it-1}\gamma_{12} + \alpha_{1,i} + \varepsilon_{1,it} \geq 0\right\} \\
y_{2,it} &= 1\left\{y_{1,it-1}\gamma_{21} + y_{2,it-1}\gamma_{22} + \alpha_{2,i} + \varepsilon_{2,it} \geq 0\right\},
\end{aligned}
$$

where $\varepsilon_{1,it}$ and $\varepsilon_{2,it}$ are logistic random variables that are independent of each other and independent over time.

All parameters are identified with $T = 4$ periods.

Relax the assumption that $\varepsilon_{1,it}$ and $\varepsilon_{2,it}$ are independent.

Following the motivation in Schmidt and Strauss (1975), Honoré and Kyriazidou (2018) consider a dynamic version of the simultaneous logit model:

$$P\left(y_{1,it} = 1 | \, y_{2,it}, y_{1,i}^{t-1}, y_{2,i}^{t-1}, \alpha_{1,i}, \alpha_{2,i} \right) = \Lambda \left(\alpha_{1,i} + y_{1,it-1}\gamma_{11} + y_{2,it-1}\gamma_{12} + \rho y_{2,it} \right)$$

and

$$P\left(y_{2,it} = 1 | \, y_{1,it}, y_{1,i}^{t-1}, y_{2,i}^{t-1}, \alpha_{1,i}, \alpha_{2,i} \right) = \Lambda \left(\alpha_{2,i} + y_{1,it-1}\gamma_{21} + y_{2,it-1}\gamma_{22} + \rho y_{1,it} \right)$$

When $\rho = 0$, this corresponds to the probabilities in Narendranthan, Nickell, and Metcalf (1985).

Honoré and Kyriazidou (2018) prove that in the dynamic model considered here, $(\gamma_{11}, \gamma_{12}, \gamma_{21}, \gamma_{22})$ is identified with at least four time periods.

However, the conditioning argument that leads to identification of $(\gamma_{11}, \gamma_{12}, \gamma_{21}, \gamma_{22})$, eliminates the parameter ρ along with the heterogeneity terms $\alpha_{1,i}$ and $\alpha_{2,i}$.

Question: What one can say about ρ for some specific data generating processes.

(Footnote: Since ρ drops out, it can be made individual-specific.)

Calculate Identified Region (Not New)

Data

$$p\left(d_2, \cdots, d_T \middle| d_1 \right)$$

Model (function of ρ)

$$\pi\left(d_2, \cdots, d_T \middle| d_1 \right) = \int \Pi_\rho\left(d_2, \cdots, d_T \middle| \alpha, d_1 \right) dH\left(\alpha \middle| y_1 = d_1 \right)$$

Approximate

$$\pi\left(d_2, \cdots, d_T \middle| d_1 \right) = \sum_k \Pi_\rho\left(d_2, \cdots, d_T \middle| \alpha_k, d_1 \right) h\left(\alpha_k \middle| y_1 = d_1 \right)$$

ρ in identified region if (for all d_1)

$$p\left(d_2, \cdots, d_T \mid d_1\right) = \sum_k \Pi_\rho\left(d_2, \cdots, d_T \mid \alpha_k, d_1\right) h\left(\alpha_k \mid y_1 = d_1\right)$$

for some $\left\{h\left(\alpha_k \mid y_1 = d_1\right)\right\}$

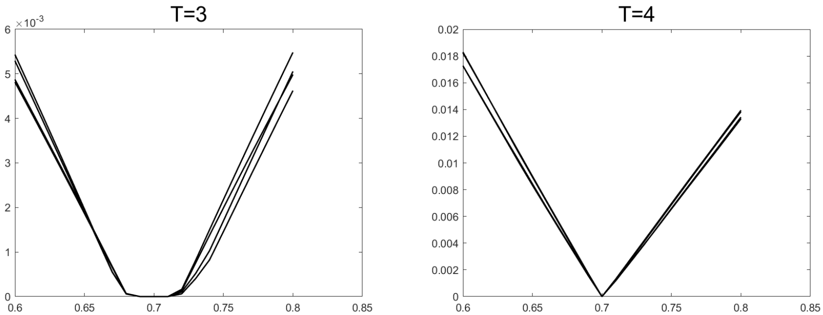
Let A_j $(j = 1, ..., 2^{T-1})$ denote all possible sequences of $d_2, \cdots, d_T$, and let

$$Q\left(g \middle| d_1\right) = \min_{\{h_k\},\{\nu_j\}} \sum_j \nu_j + \nu_0 \tag{3}$$

subject to

$$\begin{aligned}
\sum_{k=1}^{K} \Pi_g\left(A_j \middle| \alpha_k, d_1\right) h_k + \nu_j &= p\left(A_j \middle| d_1\right) \qquad (j = 1, ..., 2^{T-1}), \\
\sum_{k=1}^{K} h_k + \nu_0 &= 1, \\
\nu_j &\geq 0, \\
h_k &\geq 0.
\end{aligned}$$

Figure: Objective Functions with Negatively Correlated α's



Useful Even If You Don't Care About Fixed Effects

Paper with Anne Toft (not quite in progress) is about locational choice, but part of the paper will estimate a model of educational choice.

$$
\begin{aligned}
U_{i0} &= \delta_0^1 + x_i'\alpha + \varepsilon_i \qquad \text{(no post high school education)} \\
U_{ip} &= \delta_p^1 + x_{ip}'\alpha_p + \varepsilon_{il(p)} + \varepsilon_{is(p)} + \varepsilon_{ip} \quad \text{(program } p \text{ located at } l(p) \text{ in field } s(p)\text{)}
\end{aligned}
$$

Students rank up to 8 choices.

We will make lots of distributional assumptions

- MLE really complicated.
- Method of simulated moments conceptually easy, but difficult to get to converge.

Suppose ε_{ip} is logistic.

Consider four programs in the choice-set with $l(p_1) = l(p_2)$, $s(p_1) = s(p_2)$, $l(p_3) = l(p_4)$, and $s(p_3) = s(p_4)$. The probability that $U_{ip_1} > U_{ip_3}$ is

$$\begin{aligned} & P(U_{ip_1} > U_{ip_3}) \\ & \quad = \Lambda\left(\left(\delta^1_{p_1} + x'_{ip_1}\alpha_{p_1} - \delta^1_{p_3} - x'_{ip_3}\alpha_{p_3}\right) + \left(\varepsilon_{il(p_1)} + \varepsilon_{is(p_1)} - \varepsilon_{il(p_3)} - \varepsilon_{is(p_3)}\right)\right) \end{aligned}$$

$$\begin{aligned} & P(U_{ip_2} > U_{ip_4}) \\ & \quad = \Lambda\left(\left(\delta^1_{p_2} + x'_{ip_2}\alpha_{p2} - \delta^1_{p_4} - x'_{ip_4}\alpha_{p_4}\right) + \left(\varepsilon_{il(p_2)} + \varepsilon_{is(p_2)} - \varepsilon_{il(p_4)} - \varepsilon_{is(p_4)}\right)\right) \end{aligned}$$

These are exactly like two logits with the same fixed effect.... This will allow is to easily estimate a lot of the parameters.

Future Steps

Some parameters look like they might be point-identified, even though it seems impossible to prove it using conditioning arguments:

- Try to come up with new identification strategy
- Try to prove non–identification.
  - Chamberlain (2010) has a result for static non-logit binary outcome models
  - Working on that with Kyriazidou.
  - For the AR(1)-logit case, we seem to have that at a given $\gamma (\neq 0)$, identification is impossible outside a 2(4)-parameter family. (We think we can do better)

Link to Aguirregabiria, Gu, and Luo (2019).

-
  - Much better interpretations in terms of economics.
  - The same issues come up.