Identification and Estimation of Average Causal Effects in Fixed Effects Logit Models

Xavier D'Haultfœuille (CREST-ENSAE)

joint work with Laurent Davezies (CREST-ENSAE)
and Louise Laage (Georgetown U.)

Stata package written by Christophe Gaillac (Oxford U.)

2023 Stata Economics Virtual Symposium

9 November 2023

Outline

Binary outcomes with panels: the current practice

- Suppose we seek to identify the effect of a variable X_{kt} on a binary outcome Y_t with panel data.
- Usual parameters of interest:
  1. AME=effect on Y_T of a universal, exogenous, infinitesimal change in X_{kT}.
  2. ATE=effect on Y_T of a universal, exogenous change in X_{kT} from 0 to 1.
- Following Angrist (2001) and Angrist & Pischke (2008), applied economists most often use fixed effects (FE) linear models to estimate AME and ATE.
- Idea behind: even if wrong, such models deliver the best linear approximation of the true model.

Binary outcomes with panels: the current practice

- Yet, the results can be misleading for at least two reasons.
- 1st issue: FE linear models only use “movers” (on X); yet “stayers” may be very different from movers (and also more numerous).
- 2nd issue: nonlinearities can matter. The best linear approximation may still be bad, and identify the opposite sign of the true AME/ATE.

An alternative: the fixed effect logit model

- Logit model with fixed effects (FE):

$$\begin{aligned} Y_t &= \mathbb{1}\{X_t'\beta_0 + \alpha + \varepsilon_t \geq 0\} \\ \varepsilon_t &| \; X, \alpha \sim \text{logistic, i.i.d over } t \leq T. \end{aligned} \tag{1}$$

- "FE" approach: the distribution of $\alpha|X$ (with $X := (X_1', ..., X_T')$) is left unrestricted.
- Advantages:
  1. The model allows for heterogeneous marginal/treatment effects;
  2. The model accounts for $E(Y_t|X, \alpha) \in (0, 1)$.
- Efficient estimation of β_0 already considered by Rasch (1961); see also Andersen (1970) and Chamberlain (1980).
- But to date, no specific study of the AME and ATE in this model.

Our contribution

- We first study the identification of AME and ATE in this model:
    1. reformulate the problem as an extremal moment problem;
    2. derive simple, optimization-free, sharp bounds.
- Based on this analysis, we suggest two paths for inference:
    1. Estimate the sharp bounds.
       Requires nonparam. estimation and, for inference, some regularity on $F_{\alpha|X}$.
    2. Estimate very simple outer bounds of the AME/ATE.
       Avoids nonparam. est. and seems to work very well in practice.
- Our analysis extends to other parameters (e.g., average structural functions) and the ordered FE logit model.

Selected Literature Review

- **Marginal Effects in nonlinear FE parametric panel models**

  Honoré & Tamer (2006), Aguirregabiria and Carro (2020), Liu, Poirier and Shu (2021) ...

- **Moment problem**

  Theory: Karlin & Shapley (1953), Krein & Nudelman (1977), Schmüdgen (2017)... and old results from Chebyshev and Markov!

  Application to stats & econometrics Dette & Studden (1997), D'Haultfœuille & Rathelot (2017), Dobronyi, Gu and Kim (2021)...

Outline

The problem

- We focus on the AME at period T (say) for variable X_{kT}, defined by:

$$\Delta := E\left[\frac{\partial P(Y_T = 1|X, \alpha)}{\partial X_{kT}}\right] = \beta_{0k} E\left[\Lambda'(X_T'\beta_0 + \alpha)\right],$$

with $\Lambda(x) = 1/(1 + \exp(-x))$, $X_t = (X_{1T}, ..., X_{pT})'$ and $X = (X_1', ..., X_T')'$.

- Analysis similar for the ATE if X_{kT} is binary, and the average structural function.
- β_0 is identified by maximizing the conditional log-likelihood if

$$E\left[\sum_{s,t=1}^{T}(X_s - X_t)(X_s - X_t)'\right] \text{ is nonsingular.} \tag{2}$$

- But unclear how to get $E\left[\Lambda'(X_T'\beta_0 + \alpha)\right]$.

Intuition

- Since no constraints b/w $F_{\alpha|X=x}$ and $F_{\alpha|X=x'}$, we can focus on

$$\Delta(x) := \beta_{0k} E\left[\Lambda'(x_T'\beta_0 + \alpha)|X = x\right].$$

$\Rightarrow$ A known moment of the unobserved variable α.

- Constraints on $F_{\alpha|X=x}$, given by the data and the model.
- By sufficiency of $S = \sum_{t=1}^{T} Y_t$, all these constraints are, for $k = 0, ..., T$:

$$P(S = k|X = x) = C_k(x, \beta_0) \int \frac{\exp(ka)}{\prod_{t=1}^{T}[1 + \exp(x_t'\beta_0 + a)]} dF_{\alpha|X=x}(a)$$

where $C_k(x, \beta) = \sum_{(d_1,...,d_T)\in\{0,1\}^T : \sum_{t=1}^{T} d_t = k} \exp\left(\sum_{t=1}^{T} d_t x_t'\beta\right)$.

Intuition (c'ed)

$\Rightarrow$ For known $m, g_0, ..., g_T$, possible values of the moment $\int m(x, \alpha) dF_{\alpha|X=x}(\alpha)$, given other moments $\int g_k(x, \alpha) dF_{\alpha|X=x}(\alpha)$ $(k = 0, ..., T)$?

- A so-called moment problem.
- We first transform this moment problem into the "standard" Markov moment problem:
  1. By an appropriate transformation of the constraints;
  1. By an appropriate change of variables.
- We then use results on the Markov moment problem to solve ours.

The Markov moment problem

- Let $\mathcal{D}$ be the set of positive measures on $[0,1]$ and:

$$\mathcal{M}_T = \left\{(m_0, ..., m_T) \in \mathbb{R}^{T+1} : \exists \mu \in \mathcal{D} : \int u^t d\mu(u) = m_t,\ t = 0, ..., T\right\},$$

$$\mathcal{D}(m) = \left\{\mu \in \mathcal{D} : \int u^t d\mu(u) = m_t, t = 0, ..., T\right\} \quad \text{for } m \in \mathcal{M}_T.$$

- Then define:

$$\underline{q}_T(m) := \inf_{\mu \in \mathcal{D}(m)} \int_0^1 u^{T+1} d\mu(u),$$

$$\overline{q}_T(m) := \sup_{\mu \in \mathcal{D}(m)} \int_0^1 u^{T+1} d\mu(u).$$

- $\underline{q}_T(m)$ and $\overline{q}_T(m)$ can be obtained simply by solving univ. linear eqs Details

Some definitions

- For $t = 0, ..., T$, define:

$$\lambda_t(x, \beta) := \text{ coeff of degree } t \text{ of the polynomial}$$
$$u \mapsto u(1-u) \prod_{t=1}^{T-1} \left(1 + u\left(\exp\left((x_t - x_T)'\beta\right) - 1\right)\right),$$
$$Z_t := \binom{T-t}{S-t} \frac{\exp(SX_T'\beta_0)}{C_S(X;\beta_0)},$$
$$m_t(x) := \frac{E(Z_t|X=x)}{E(Z_0|X=x)},$$
$$m(x) := (m_0(x), ..., m_T(x))'.$$

- To remember here: all these are identified and easy to estimate, except $m(x)$ that involves conditional expectations.

Key result

Theorem 1

Suppose that (1)-(2) hold. Then, there exists a collection of probability measures $(\mu_x)_{x\in Supp(X)}$*, with* $\mu_x \in \mathcal{D}(m(x))$*, such that*

$$\Delta = \beta_{0k} E\left[\sum_{t=0}^{T} Z_t \lambda_t(x;\beta_0) + Z_0 \lambda_{T+1}(X,\beta_0) \int_0^1 u^{T+1} d\mu_X(u)\right]. \quad (3)$$

Moreover, the sharp identified set of Δ *is* $[\underline{\Delta}, \overline{\Delta}]$*, with*

$$\begin{aligned}\underline{\Delta} =& E\left[\sum_{t=0}^{T} Z_t \lambda_t(x;\beta_0) + \beta_{0k} Z_0 \lambda_{T+1}(X,\beta_0)\big(\underline{q}_T(m(X))\right.\\ & \left.\mathbb{1}\left\{\beta_{0k}\lambda_{T+1}(X,\beta_0)\geq 0\right\} + \overline{q}_T(m(X))\mathbb{1}\left\{\beta_{0k}\lambda_{T+1}(X,\beta_0) < 0\right\}\big)\right],\\ \overline{\Delta} =& E\left[\sum_{t=0}^{T} Z_t \lambda_t(x;\beta_0) + \beta_{0k} Z_0 \lambda_{T+1}(X,\beta_0)\big(\overline{q}_T(m(X))\right.\\ & \left.\mathbb{1}\left\{\beta_{0k}\lambda_{T+1}(X,\beta_0)\geq 0\right\} + \underline{q}_T(m(X))\mathbb{1}\left\{\beta_{0k}\lambda_{T+1}(X,\beta_0) < 0\right\}\big)\right].\end{aligned}$$

Simple outer bounds: idea

- Drawback of the sharp bounds: use $m(x)$, which requires nonparam. estim.
- Actually, Eq. (3) also useful for obtaining simple outer bounds.
- Δ is not identified solely because of $\int_0^1 u^{T+1} d\mu_X(u)$.
- Imagine that instead of u^{T+1}, we had $P(u) = \sum_{k=0}^{T} b_k u^k$.
- Then, using $\int_0^1 u^k d\mu_X(u) = E(Z_k|X)/E(Z_0|X)$, we would get for Δ:

$$\beta_{0k} E\left[\sum_{t=0}^{T} \left(\lambda_t(X, \beta_0) + b_t \lambda_{T+1}(X, \beta_0)\right) Z_t\right].$$

Very simple expectation!

Simple outer bounds: idea (c'ed)

- Now, if $\sup_{u\in[0,1]}\left|u^{T+1}-\sum_{k=0}^{T} b_k u^k\right| \leq K$ for some $K>0$, we obtain the outer bounds for Δ:

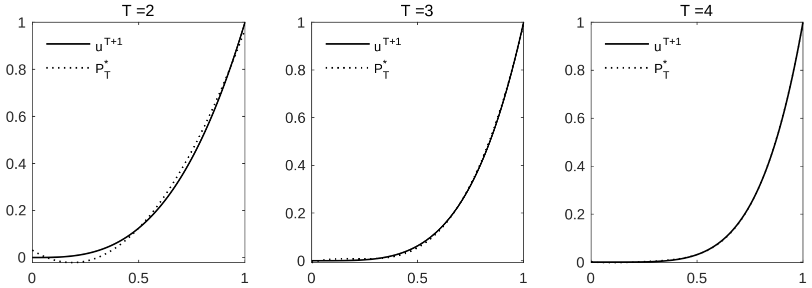
$$\left[\beta_{0k} E\left(\sum_{t=0}^{T}\left(\lambda_t(X,\beta_0)+b_t\lambda_{T+1}(X,\beta_0)\right)Z_t\right) \pm K\, E\left(Z_0\left|\beta_{0k}\lambda_{T+1}(X,\beta_0)\right|\right)\right].$$

- We can optimize these bounds, by choosing appropriately $(b_0,...,b_T)$.
- Specifically, we consider the best sup-norm approximation of $u\mapsto u^{T+1}$ by a polynomial of degree T:

$$\mathbf{b}^* = \underset{\mathbf{b}\in\mathbb{R}^{T+1}}{\text{argmin}} \sup_{u\in[0,1]}\left|u^{T+1}-\sum_{k=0}^{T} b_k u^k\right| \quad (4)$$

Simple outer bounds

- $\mathbf{b}^*$ very simple to compute, using Chebyshev polynomials.
- Figures below plot $u \mapsto u^{T+1}$ and $P_T^*(u) = \sum_{k=0}^{T} b_t^* u^t$.



- P_2^* approximates already very well $u \mapsto u^3$, curves indistinguishable for $T = 4$.
- With $\mathbf{b} = \mathbf{b}^*$, we have $K = 1/(2 \times 4^T)$.

Are the bounds informative?

Proposition 1 (Some properties of the bounds on Δ)

Suppose that (1)-(2) *hold. Then:*

1. *The outer bounds may coincide with the sharp bounds.*
2. $\overline{\Delta} - \underline{\Delta} \leq E\left[Z_0 | \lambda_{T+1}(X, \beta_0)|\right]/4^T$. *If also* $|(X_t - X_T)'\beta_0| \leq \ln(2)$ *a.s.,*
$$\overline{\Delta} - \underline{\Delta} \leq \frac{1}{4^T}.$$
3. Δ *is point identified if and only if* $\beta_{0k} = 0$ *or*
$$P\left(\min_{t<T} |(X_t - X_T)'\beta_0| = 0 \cup |Supp(\alpha|X)| \leq T/2\right) = 1.$$

Outline

Estimation of the sharp bounds

- Recall that

$$\overline{\Delta} = E\left[\sum_{t=0}^{T} Z_t \lambda_t(x;\beta_0) + \beta_{0k} Z_0 \lambda_{T+1}(X,\beta_0)\big(\overline{q}_T(m(X)) \right.$$
$$\left. \mathbb{1}\left\{\beta_{0k}\lambda_{T+1}(X,\beta_0) \geq 0\right\} + \underline{q}_T(m(X))\mathbb{1}\left\{\beta_{0k}\lambda_{T+1}(X,\beta_0) < 0\right\}\big)\right],$$
$$m(X) = (m_0(X), ..., m_T(X))', \quad m_t(x) := \frac{E(Z_t|X=x)}{E(Z_0|X=x)},$$
$$Z_t = \binom{T-t}{S-t}\frac{\exp(SX_T'\beta_0)}{C_S(X;\beta_0)}.$$

- All terms can be estimated easily, except $m_t(X)$.
- We first estimate by local polynomial regression $E(Z_t|X=x)$ and obtain a plug-in estimator of $m(X)$.
- We modify this initial estimator to ensure that $\widehat{m}(X)$ is a true moment vector (e.g., the corresponding variance is positive).

Asymptotic distribution of $(\widehat{\underline{\Delta}}, \widehat{\overline{\Delta}})$

Theorem 2

Suppose we have i.i.d. data and (1)-(2) and Assumption 1 hold Details*. Then, there exist* $(\underline{\psi}_i, \overline{\psi}_i)_{i=1,\dots,n}$ *i.i.d. such that:*

1. *If* $\beta_{0k} > 0$, *then*

$$\sqrt{n}\begin{pmatrix} \widehat{\overline{\Delta}} - \overline{\Delta} \\ \widehat{\underline{\Delta}} - \underline{\Delta} \end{pmatrix} = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\begin{pmatrix} \overline{\psi}_i \\ \underline{\psi}_i \end{pmatrix} + o_P(1).$$

   If $\beta_{0k} < 0$, *same but with* $\underline{\psi}_i$ *and* $\overline{\psi}_i$ *switched.*

2. *If* $\beta_{0k} = 0$, *then*

$$\sqrt{n}\begin{pmatrix} \widehat{\overline{\Delta}} - \overline{\Delta} \\ \widehat{\underline{\Delta}} - \underline{\Delta} \end{pmatrix} = \begin{pmatrix} \max\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\overline{\psi}_i, \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\underline{\psi}_i\right) \\ \min\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\overline{\psi}_i, \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\underline{\psi}_i\right) \end{pmatrix} + o_P(1).$$

- We also show that we can consistently estimate $\Sigma := V((\underline{\psi}, \overline{\psi})')$.

Construction of confidence intervals (CI)

- The estimated bounds are not asymptotically normal when $\beta_{0k} = 0$.

$\Rightarrow$ The CI of Imbens and Manski (2004) works when $\beta_{0k} \neq 0$, but possibly not when $\beta_{0k} = 0$.

- We modify them in a simple way. Let φ_α a test of $\beta_{0k} = 0$. Then let:

$$\text{CI}^1_{1-\alpha} := \begin{vmatrix} \left[\underline{\widehat{\Delta}} - c_\alpha \left(\frac{\widehat{\Sigma}_{11}}{n}\right)^{1/2}, \overline{\widehat{\Delta}} + c_\alpha \left(\frac{\widehat{\Sigma}_{22}}{n}\right)^{1/2}\right] & \text{if } \varphi_\alpha = 1, \\ \left[\min\left(0, \underline{\widehat{\Delta}} - c_\alpha \left(\frac{\widehat{\Sigma}_{11}}{n}\right)^{1/2}\right), \max\left(0, \overline{\widehat{\Delta}} + c_\alpha \left(\frac{\widehat{\Sigma}_{22}}{n}\right)^{1/2}\right)\right] & \text{if } \varphi_\alpha = 0. \end{vmatrix}$$

where c_α is defined as in I & M.

Proposition 2

Suppose we have i.i.d. data, (1)-(2) *and A1 hold and* $\min(\Sigma_{11}, \Sigma_{22}) > 0$. *Then* $\liminf_n \inf_{\Delta \in [\underline{\Delta}, \overline{\Delta}]} P(\Delta \in CI^1_{1-\alpha}) \geq 1 - \alpha$*, with equality when* $\beta_{0k} \neq 0$.

Inference using outer bounds

- The outer bounds take the form $[\widetilde{\Delta} \pm \overline{b}]$, with

$$\tilde{\Delta} = E\left[\sum_{t=0}^{T} Z_t \left(\lambda_t(X, \beta_0) + b^*_{t,T}\lambda_{T+1}(X, \beta_0)\right)\right],$$

$$\overline{b} = \frac{1}{2 \times 4^T} E\left[Z_0 |\lambda_{T+1}(X, \beta_0)|\right].$$

- We can estimate these simply by plug-in $\Rightarrow \widehat{\tilde{\Delta}}$ and $\widehat{\overline{b}}$.
- We then consider the confidence interval

$$\text{CI}^2_{1-\alpha} = \left[\widehat{\tilde{\Delta}} \pm q_\alpha\left(\frac{n^{1/2}\widehat{\overline{b}}}{\widehat{\sigma}}\right)\frac{\widehat{\sigma}}{n^{1/2}}\right],$$

where $q_\alpha(b)$ = quantile of order $1 - \alpha$ of a $|\mathcal{N}(b, 1)|$ and $\widehat{\sigma}$ is an estimator of the asymptotic variance of $\widehat{\tilde{\Delta}}$.

Construction of confidence intervals on Δ

Theorem 3

Suppose (1)-(2) hold, X is bounded and either $|\tilde{\Delta} - \Delta| < \overline{b}$ or $\beta_{0k} = 0$. Then:

$$\liminf_{n\to\infty} P\left(\Delta \in CI^2_{1-\alpha}\right) \geq 1 - \alpha.$$

- $|\tilde{\Delta} - \Delta| < \overline{b}$ or $\beta_{0k} = 0$ holds except if

$$P\left(\text{Supp}(\Lambda(X_T'\beta_0 + \alpha)|X) \subset \mathcal{R}_X | \lambda_{T+1}(X, \beta_0) \neq 0\right) = 1, \quad (5)$$

  where $\mathcal{R}_x$ is the set of maxima (resp. minima) of the polynomial $\mathbb{T}_{T+1}$ on $[0,1]$ if $\lambda_{T+1}(x, \beta_0) > 0$ (resp. $\lambda_{T+1}(x, \beta_0) < 0$).
- Eq. (5) unlikely: it implies a very specific location for $\text{Supp}(\alpha)|X = x)$, with discontinuous changes in this support at some x.
- But $\text{CI}^2_{1-\alpha}$ may not have a uniform coverage. See the paper for a slightly larger CI, uniform over a large class of DGP.

Outline

Designs

- We assume $X_1, ..., X_T$ i.i.d., with $X_t \in \mathbb{R} \sim \mathcal{U}[-1/2, 1/2]$ and $\beta_0 = 1$.
- We let $T \in \{2, 3\}$ and $n \in \{250; 500; 1,000\}$.
- We then let $\alpha = -X_T'\beta_0 + \eta$, with either:
  1. $\eta | X \sim \mathcal{N}(0, 1)$;
  2. or $\eta | X$ such that $\tilde{\Delta} - \Delta = \overline{b}$.
- In the 2nd case, the DGP varies with T.

DGP1: estimators of the bounds

- $\Delta \simeq 0.2067$ is partially identified for all T.
- $(\underline{\Delta}, \overline{\Delta}) \simeq (0.2006, 0.2124)$ if $T = 2$ and $(\underline{\Delta}, \overline{\Delta}) \simeq (0.2059, 0.2069)$ if $T = 3$.

		First method				Second method	
T	n	$\sigma(\widehat{\underline{\Delta}})$	Bias$(\widehat{\underline{\Delta}})$	$\sigma(\widehat{\overline{\Delta}})$	Bias$(\widehat{\overline{\Delta}})$	$\sigma(\widehat{\tilde{\Delta}})$	Bias$(\widehat{\tilde{\Delta}})$
2	250	0.110	0.006	0.114	0.003	0.108	0.002
	500	0.077	0.013	0.081	0.01	0.074	0.005
	1000	0.054	0.013	0.057	0.011	0.052	0.004
3	250	0.072	-0.005	0.072	-0.005	0.074	-0.001
	500	0.049	-0.003	0.049	-0.004	0.051	0*
	1000	0.035	-0.004	0.036	-0.005	0.037	-0.001

Notes: *: absolute value < 0.0005. Results obtained with 3,000 sims.

DGP1: comparison between the two CI's

		$CI^1_{0.95}$		$CI^2_{0.95}$	
T	n	coverage	av. length	coverage	av. length
2	250	0.96	0.453	0.96	0.419
	500	0.96	0.305	0.96	0.296
	1000	0.95	0.215	0.97	0.211
3	250	0.96	0.288	0.95	0.284
	500	0.96	0.201	0.95	0.201
	1000	0.95	0.141	0.94	0.142

Notes: results obtained with 3,000 sims.

DGP2$_T$: estimators of the bounds

- $\Delta = \underline{\Delta} = \overline{\Delta} \simeq 0.1875$ if $T = 2$
  $\Delta \simeq 0.1667$ and $(\underline{\Delta}, \overline{\Delta}) \simeq (0.1652, 0.1667)$ if $T = 3$.

		First method				Second method	
T	n	$\sigma(\widehat{\underline{\Delta}})$	Bias($\widehat{\underline{\Delta}}$)	$\sigma(\widehat{\overline{\Delta}})$	Bias($\widehat{\overline{\Delta}}$)	$\sigma(\widehat{\tilde{\Delta}})$	Bias($\widehat{\tilde{\Delta}}$)
2	250	0.146	0.049	0.151	0.058	0.105	-0.003
	500	0.104	0.032	0.108	0.041	0.076	-0.009
	1000	0.069	0.026	0.072	0.034	0.052	-0.01
3	250	0.075	0.01	0.075	0.009	0.063	0.001
	500	0.05	0.005	0.05	0.004	0.045	-0.001
	1000	0.034	0.005	0.034	0.004	0.031	0*

Notes: *: abs. value < 0.0005. Results obtained with 3,000 sims.

- The biases of $(\widehat{\underline{\Delta}}, \widehat{\overline{\Delta}})$ are not that small when $n = 1,000$ and $T = 2$.
- This could be b/c regularity conditions on $\gamma_s(.)$ are actually violated here.

DGP2$_T$: comparison between the two CI's

		$\text{CI}^1_{0.95}$		$\text{CI}^2_{0.95}$	
T	n	coverage	av. length	coverage	av. length
2	250	0.92	0.522	0.96	0.420
	500	0.91	0.353	0.95	0.295
	1000	0.91	0.243	0.95	0.209
3	250	0.96	0.276	0.96	0.249
	500	0.95	0.186	0.95	0.175
	1000	0.95	0.130	0.95	0.124

Notes: results obtained with 3,000 sims.

- $\text{CI}^2_{1-\alpha}$ has still very good coverage, though $\tilde{\Delta} - \Delta = \overline{b}$.
- $\text{CI}^1_{1-\alpha}$ undercovers for $T = 2$, probably b/c of the aforementioned bias.

Outline

What it does & does not do yet

- Available on SSC (requires `estout` to be installed).
- First estimates β_0, then computes estimated bounds or the "point estimate" $\widehat{\tilde{\Delta}}$, and CI for the AME (if not binary) or the ATE (if binary).
- Handles unbalanced panels.
- Still in progress:
  - Does not handle factor variables yet;
  - Only estimates the sharp bounds & $\text{CI}^1_{1-\alpha}$ for continuous X;
  - Does not handle, e.g. age and age^2;
  - Could probably be faster.

Simplified syntax

mfelogit *depvar* [*indepvar*] [*if*] [*in*] method(*string*) id(*string*) time(*string*) [, listT(*string*) listX(*string*) level(*string*)]

- id and time: individual and time identifiers.
- method: outer bounds if "quick" (default), sharp bounds if "sharp".
- listT: periods on which AME / ATE are computed. By default, last period for which all individuals are observed. If "all", computes AME / ATE for all periods, and their averages.
- listX: covariates for which the AME / ATE are computed. By default, all covariates.
- level: level of confidence intervals. "0.95" by default.

(Toy) example: determinants of unionization in the US

- Syntax:

```
use "https://www.stata-press.com/data/r17/union.dta", clear

tabulate year, generate(y_)
drop y_1

mfelogit union south y_* black, id("idcode") time("year")

xtset idcode year
xtreg union age y_* black south, fe
```

- `black` automatically omitted as constant for each indiv. over time.
- Results on the ATE for `south`, with the FE logit and FE linear regs.:

	FE Logit model	FE linear reg.
Point est.	-.072	-.071
95% CI	[-.095, -.048]	[-.103, -.040]

Outline

Conclusion

- Simple characterization of the identified set for the AME.
- Based on this, estimators of the sharp bounds of the AME.
- Alternative method based on a "proxy" of the AME and an upper bound on its asymptotic bias.
- Though not optimal as $n \to \infty$, very simple and seems to work very well for usual n and T.
- Already a Stata command, `mfelogit`, available on SSC. Will be improved soon hopefully!

Intuition on the Markov moment problem for $T = 1$.

- If $T = 1$, we seek bounds on $\int_0^1 x^2 d\mu(x)$ given $\int_0^1 x d\mu(x) = m_1$.
- Using $x^2 \leq x$ on $[0, 1]$ and Jensen's ineq., we get $\underline{q}_1(m) = m_1^2$, $\overline{q}_1(m) = m_1$.

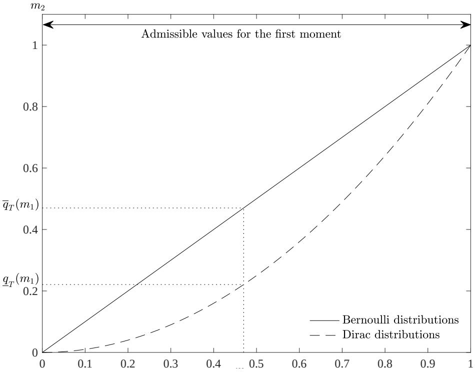


Figure: Moment space and bounds $\underline{q}_T(m), \overline{q}_T(m)$ when $T = 1$.

Solving the moment problem for any T Back

- Such ideas generalize for any T: $\underline{q}_T(m)$ and $\overline{q}_T(m)$ rational functions of m.
- Let $T > 0$ and for any $m = (m_0, ..., m_s)$, $s > T$, let

$$\begin{array}{lll} \underline{\mathbb{H}}_T(m) = (m_{i+j-2})_{1\le i,j\le T/2+1}, & \overline{\mathbb{H}}_T(m) = (m_{i+j-1} - m_{i+j})_{1\le i,j\le T/2} & \text{if } T \text{ even} \\ \underline{\mathbb{H}}_T(m) = (m_{i+j-1})_{1\le i,j\le (T+1)/2}, & \overline{\mathbb{H}}_T(m) = (m_{i+j-2} - m_{i+j-1})_{1\le i,j\le (T+1)/2} & \text{if } T \text{ odd.} \end{array}$$

- Then let $\underline{H}_T(c) = det\left(\underline{\mathbb{H}}_T(c)\right)$ and $\overline{H}_T(c) = det\left(\overline{\mathbb{H}}_T(c)\right)$.

Proposition 3 (Extremal moments & Hankel determinants)

1. $\mathcal{M}_T = closure\left\{m \in \mathbb{R}^{T+1} : \underline{H}_t(m) > 0 \text{ and } \overline{H}_t(m) > 0, t = 1, ..., T\right\}$.
2. *If* $m \in \mathcal{M}_T$ *and* $\underline{H}_T(m) \times \overline{H}_T(m) > 0$, $\underline{q}_T(m) < \overline{q}_T(m)$. *Also,* $q \mapsto \underline{H}_{T+1}(m, q)$ *is strictly* $\uparrow$*, linear and*

$$\underline{H}_{T+1}(m, \underline{q}_T(m)) = 0 \quad (\text{and similarly for } \overline{q}_T(m)).$$

- See the paper for the point identified case (when $\underline{H}_T(m) \times \overline{H}_T(m) = 0$).

Conditions for asymptotic normality in the 1st method

Back

- Let $\gamma(.) = (\gamma_0(.), ..., \gamma_T(.))$ with $\gamma_t(x) = P(S = t|X = x)$.
- K be the kernel in the local polynomial (of degree $\ell \geq pT/2$) estimator of $\gamma_t(.)$ and $h_n \in \mathbb{R}$ be the bandwidth.

Assumption 1

1. *K has a compact support and is Lipschitz on $\mathbb{R}^{pT}$.*
2. $nh_n^{2(\ell+1)} \to 0$ *and* $n[h_n^{pT}/\ln n]^3 \to \infty$.
3. *The pdf of X, f_X, is C^1 and bounded away from 0 on its bounded support.*
4. γ_0 *is* $C^{\ell+2}$ *on* $Supp(X)$.
5. *Either* $|Supp(\alpha|X = x)| > T/2$ *for all* $x \in Supp(X)$, *or* $x \mapsto |Supp(\alpha|X = x)$ *is constant.*

- Point 5 needed b/c $\underline{q}_t$ and $\overline{q}_t$ not differentiable at all $m \in \partial\mathcal{M}_t$ if $t \geq 3$.