MarginalFElogit:
R Package for the Estimation of Average Causal Effects in Fixed Effect Logit Models

Laurent Davezies CREST-ENSAE

Xavier D'Haultfœuille CREST-ENSAE

Christophe Gaillac University of Geneva

Louise Laage Georgetown University

Maël Laoufi CREST-ENSAE

Abstract

This vignette presents the package **MarginalFElogit**, which estimates bounds and performs inference on average causal effects in a panel data fixed effects binary logit model. It computes the estimators of the sharp and outer bounds on these parameters derived in Davezies et al. (2024), as well as the corresponding confidence intervals. We illustrate the usage of **MarginalFElogit** with a simulated and a real example.

Keywords: Fixed effects logit models, panel data, partial identification, R.

---

1. How to get started

R is an open source software project and can be freely downloaded from the CRAN website (`http://cran.r-project.org`). The R package **MarginalFElogit** can be downloaded from Github (`https://github.com/cgaillac/MarginalFElogit`). To install the **MarginalFElogit** package from Github, the devtools library is required. Then, use the command

```
library("devtools")
install_github('cgaillac/MarginalFElogit')
```

Online help is available in two ways: either help(package="MarginalFElogit") or ?felogit. The first gives an overview over the available commands in the package. The second gives detailed information about a specific command. A valuable feature of R help files is that the examples used to illustrate commands are executable, so they can be pasted into an R session or run as a group with a command like example(felogit).

2. Theory

Consider a panel with T periods where for each individual $i = 1, ..., n$ and each period $t = 1, ..., T$, we observe a binary outcome Y_{it} and a vector of covariates $X_{it} := (X_{it1}, ..., X_{itp})'$. Let $X_i = (X'_{i1}, ..., X'_{iT})'$. We consider the logit model with fixed effects:

$$\begin{aligned} Y_{it} &= \mathbb{1}\left\{X'_{it}\beta_0 + \alpha_i + \varepsilon_{it} \geq 0\right\}, t = 1, ..., T, \\ \varepsilon_{it} &\mid X_i, \alpha \sim \text{logistic, i.i.d over } t. \end{aligned} \tag{1}$$

Importantly, the individual effect α_i is allowed to be correlated in an unspecified way with covariates X_i. For simplicity, we first omit index i below. Davezies, D'Haultfœuille, and Laage (2024, DDL hereafter) develop methods to compute the bounds on average causal effects, including the following two standard parameters:

1. The Average Marginal Effect (AME) at period t (AME_t), namely the effect on Y_t of a universal, exogenous, infinitesimal change in X_{kt}. Using (1), we have $P(Y_t = 1|X, \alpha) = \Lambda(X_t'\beta_0 + \alpha)$ with $\Lambda(x) := 1/(1 + \exp(-x))$. Hence,

$$\text{AME}_t := \mathsf{E}\left[\frac{\partial P(Y_t = 1|X, \alpha)}{\partial X_{tk}}\right] = \beta_{0k}\mathsf{E}[\Lambda'(X_t'\beta_0 + \alpha)],$$

where $\beta_0 = (\beta_{01}, ..., \beta_{0p})'$.

2. The Average Treatment Effect (ATE) at period t (ATE_t), namely the effect on Y_t of a universal, exogenous change in X_{kt} from 0 to 1,

$$\text{ATE}_t := \mathsf{E}\left[\Lambda(X_{t-k}'\beta_{0-k} + \beta_{0k}) - \Lambda(X_{t-k}'\beta_{0-k})\right],$$

where X_{t-k} (resp. β_{0-k}) denotes the vector X_t (resp. β_0) without its k-th component.

As shown by DDL, the AME_t and parameters are partially identified in general. We explain briefly below how DDL obtain the sharp and some outer bounds on them. First, recall that because $S := \sum_{t=1}^T Y_t$ is a sufficient statistic for α (e.g. Andersen 1970), β_0 can be identified by maximizing the expected conditional log-likelihood, conditioning not only on X but also on S.

Next, let $v(x, \beta) := x_t'\beta$ for the AME_t and $v(x, \beta) := x_t'\beta - (2x_{tk} - 1)\beta_k$ for the ATE_t. Let also $\Omega_{x,\beta}(u) := \prod_{t=1}^T [u(\exp(x_t'\beta - v(x, \beta)) - 1) + 1]$, $\lambda_t(x, \beta)$ ($t = 0, ..., T+1$) denote the coefficient of u^t in the polynomial $u \mapsto \beta_k u(1 - u)\Omega_{x,\beta}(u)$ for the AME_t and $u \mapsto -(2x_{tk} - 1)u\Omega_{x,\beta}(u)$ for the ATE_t. Let also

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

$$Z_t(x, s, \beta) := \binom{T-t}{s-t}\frac{\exp(sv(x, \beta))}{C_s(x; \beta)}, \quad t = 0, ..., T.$$

Finally, let $Z_t := Z_t(X, S, \beta_0)$. Then, DDL show that

$$\text{AME}_t = \mathsf{E}\left[\sum_{t=0}^T \lambda_t(X, \beta_0)Z_t + \lambda_{T+1}(x, \beta_0)Z_0 \int_0^1 u^{T+1}d\mu_X(u)\right], \tag{2}$$

$$\text{ATE}_t = \mathsf{E}\left[(2X_{tk} - 1)Y_t + \sum_{t=0}^T \lambda_t(X, \beta_0)Z_t + \lambda_{T+1}(x, \beta_0)Z_0 \int_0^1 u^{T+1}d\mu_X(u)\right], \tag{3}$$

where μ_x is a probability measure on $[0, 1]$ satisfying $\mu_x(\{0, 1\}) = 0$ and

$$\int_0^1 u^k \mu_x(u)du = m_t(x) := \frac{\mathsf{E}[Z_t|X = x]}{\mathsf{E}[Z_0|X = x]}.$$

In (2) and (3), only $m_{T+1}(X) := \int_0^1 u^{T+1} d\mu_X(u)$ is not point identified in general. The idea, then, is to find either sharp or outer bounds on this quantity to then bound AME_t and ATE_t, using (2) and (3).

2.1. Sharp bounds: identification, estimation and inference

Identification

The sharp bounds on $m_{T+1}(X)$ are obtained using results from the theory of moments (see, e.g., Schmüdgen 2017, for a mathematical exposition). We require additional notation. For any $t \geq 1$, $s \geq t$ and $m = (m_0, ..., m_s) \in \mathbb{R}^{s+1}$, we define the Hankel matrices $\underline{\mathbb{H}}_t(m)$ and $\overline{\mathbb{H}}_t(m)$ as

$$\begin{aligned} &\underline{\mathbb{H}}_t(m) = (m_{i+j-2})_{1 \leq i,j \leq t/2+1}, && \overline{\mathbb{H}}_t(m) = (m_{i+j-1} - m_{i+j})_{1 \leq i,j \leq t/2} && \text{if } t \text{ is even,} \\ &\underline{\mathbb{H}}_t(m) = (m_{i+j-1})_{1 \leq i,j \leq (t+1)/2}, && \overline{\mathbb{H}}_t(m) = (m_{i+j-2} - m_{i+j-1})_{1 \leq i,j \leq (t+1)/2} && \text{if } t \text{ is odd.} \end{aligned}$$

Then, let $\underline{H}_t(m) = \det(\underline{\mathbb{H}}_t(m))$ and $\overline{H}_t(m) = \det(\overline{\mathbb{H}}_t(m))$. Now, for $m \in \mathbb{R}^{T+1}$ and $q \in \mathbb{R}$, consider $\underline{H}_{T+1}(m, q)$. By expanding this determinant along its last column, we see that $q \mapsto \underline{H}_{T+1}(m, q)$ is linear. Then, let $\underline{a}_{T+1}(m)$ and $\underline{b}_{T+1}(m)$ denote the corresponding intercept and slope parameters, so that $\underline{H}_{T+1}(m, q) = \underline{a}_{T+1}(m) + \underline{b}_{T+1}(m)q$. We define similarly $\overline{a}_{T+1}(m)$ and $\overline{b}_{T+1}(m)$. The sharp lower and upper bounds on $m_{T+1}(X)$ are then respectively $\underline{q}_T(m(X))$ and $\overline{q}_T(m(X))$, where $m(X) = (m_0(X), ..., m_T(X))$ and the functions $\underline{q}_T$ and $\overline{q}_T$ are defined as follows:

1. If $\underline{H}_T(m) \times \overline{H}_T(m) > 0$, then $\underline{q}_T(m) = -\underline{a}_{T+1}(m)/ \underline{b}_{T+1}(m)$ and $\overline{q}_T(m) = -\overline{a}_{T+1}(m) /\overline{b}_{T+1}(m)$.
2. If $\underline{H}_T(m) \times \overline{H}_T(m) = 0$, then $\underline{q}_T(m)$ is equal to

$$\begin{aligned} &-\underline{a}_{T'}(m_{T-T'+1}, ..., m_T)/\underline{b}_{T'}(m_{T-T'+1}, ..., m_T) \text{ if } \underline{H}_{T'}(m) = 0, \\ &-\overline{a}_{T'}(m_{T-T'+1}, ..., m_T)/\overline{b}_{T'}(m_{T-T'+1}, ..., m_T) \text{ if } \overline{H}_{T'}(m) = 0, \end{aligned}$$

   where $T' = \min\{t \leq T : \underline{H}_t(m) \times \overline{H}_t(m) = 0\}$. Moreover, $\underline{q}_T(m) = \overline{q}_T(m)$.

With this definition of $\underline{q}_T$ and $\overline{q}_T$ and (2), we finally obtain the sharp lower and upper bounds ($\underline{\text{AME}}_t$ and $\overline{\text{AME}}_t$) on AME_t:

$$\begin{aligned} \underline{\text{AME}}_t = \mathsf{E}\Big[\sum_{t=0}^{T} \lambda_t(X, \beta_0) Z_t + \lambda_{T+1}(x, \beta_0) Z_0 \Big(& \mathbb{1}\left\{\lambda_{T+1}(x, \beta_0 > 0\right\} \underline{q}_T(m(X)) \\ & + \mathbb{1}\left\{\lambda_{T+1}(x, \beta_0 < 0\right\} \overline{q}_T(m(X)) \Big) \Big], \\ \overline{\text{AME}}_t = \mathsf{E}\Big[\sum_{t=0}^{T} \lambda_t(X, \beta_0) Z_t + \lambda_{T+1}(x, \beta_0) Z_0 \Big(& \mathbb{1}\left\{\lambda_{T+1}(x, \beta_0 > 0\right\} \overline{q}_T(m(X)) \\ & + \mathbb{1}\left\{\lambda_{T+1}(x, \beta_0 < 0\right\} \underline{q}_T(m(X)) \Big) \Big]. \end{aligned}$$

The sharp bounds on the ATE_t are obtained similarly, using (3).

Estimation

We now suppose to observe a sample $(Y_i, X_i)_{i=1,...,n}$ of i.i.d. variables with the same distribution as (Y, X), with $Y := (Y_1, ..., Y_T)$. We estimate $\underline{\text{AME}}_t$ and $\overline{\text{AME}}_t$ in four steps.Note that the third step is necessary to ensure that the estimator of $m(X)$ is indeed a vector of moment of a probability measure on $[0,1]$ (the functions $\underline{q}_T$ and $\overline{q}_T$ are only defined on such vectors):

1. Estimation of β_0:

   Let us define the conditional likelihood as

$$\ell_c(y|x;\beta) := \sum_{t=1}^T y_t x_t'\beta - \ln\left[C_{\sum_{t=1}^T y_t}(x,\beta)\right],$$

   for a vector $y = (y_1, ..., y_T)'$ and $x = (x_1', ..., x_T')' \in \mathbb{R}^{pT}$. Then, define the conditional maximum likelihood estimator as

$$\widehat{\beta} := \arg\max_{\beta\in\mathbb{R}^p} \sum_{i=1}^n \ell_c(Y_i|X_i;\beta).$$

2. Initial estimation of $m = (m_0, ..., m_T)$:

   Let $\gamma_{0j}(x) := P(S = j|X = x)$ for $j = 0, ..., T$. By definition, $m_t(x) = c_t(x)/c_0(x)$, with

$$c_t(x) := \mathsf{E}(Z_t|X = x) = \sum_{j=t}^T \binom{T-t}{j-t} \frac{\gamma_{0j}(x)\exp(jv(x,\beta_0))}{C_j(x,\beta_0)}. \tag{4}$$

   We first estimate the $(\gamma_{0j})_{j=0,...,T}$ nonparametrically. To avoid boundary effects, we consider a local polynomial estimator $\widehat{\gamma}_j$ of order ℓ. For any $(b,x) \in \mathbb{N}^{pT} \times \mathbb{R}^{pT}$, let $|b| := \sum_{j=1}^{pT} b_j$ and $x^b := x_1^{b_1}...x_{pT}^{b_{pT}}$. Let $\widehat{a}^j(x) = (\widehat{a}_b^j(x))_{b\in\mathbb{N}^{pt}:|b|\leq\ell}$ be defined as

$$\widehat{a}^j(x) := \text{argmin}_{a\in\mathbb{R}^{tp}} \sum_{i=1}^n K\left(\frac{X_i - x}{h_n}\right)\left(\mathbb{1}\{S_i = j\} - \sum_{b\in\mathbb{N}^{pt}:|b|\leq\ell} a_b (X_i - x)^b\right)^2, \tag{5}$$

   where $S_i := \sum_{t=1}^T Y_{it}$, K denotes the kernel function and h_n is the bandwidth. The estimator of $\gamma_{0j}(x)$ is $\widehat{\gamma}_j(x) := \widehat{a}_0^j(x)$.

   Then, let $\widehat{c}_t(x)$ be the plug-in estimator of $c_t(x)$ based on (4), replacing β_0 and $\gamma_{0j}(x)$ by respectively $\widehat{\beta}$ and $\widehat{\gamma}_j(x)$. Finally, let

$$\widetilde{m}_t(x) := \widehat{c}_t(x)/\widehat{c}_0(x), \quad t = 0, ..., T.$$

3. Constrained estimation of m:

   For any $(m_t)_{t\geq 0}$ and $t \in \{0, ..., T\}$, let $m_{\to t} := (m_0, ..., m_t)$. Let c_n be a sequence tending to 0 (we use c_n proportional to $[\ln(\ln(n))/n]^{1/2}$ in practice) and define

$$\widehat{I}(x) := \max\left\{t \in \{1, ..., T\} : \forall s \leq t,\ \underline{H}_s(\widetilde{m}_{\to s}(x)) \times \overline{H}_s(\widetilde{m}_{\to s}(x)) > c_n\right\}, \tag{6}$$

with the convention that $\max \emptyset = 0$. We then let $\widehat{m}_{\to \widehat{I}(x)}(x) := \widetilde{m}_{\to \widehat{I}(x)}(x)$. If $\widehat{I}(x) = T$, $\widehat{m}(x)$ is fully defined. Otherwise, we complete $\widehat{m}(x)$ by first letting

$$\widehat{m}_{\widehat{I}(x)+1}(x) := \left| \begin{array}{ll} \underline{q}_{\widehat{I}(x)}(\widetilde{m}_{\to \widehat{I}(x)}(x)) & \text{if } \underline{H}_{\widehat{I}(x)+1}(\widetilde{m}_{\to \widehat{I}(x)+1}(x)) < c_n^{1/2}, \\ \overline{q}_{\widehat{I}(x)}(\widetilde{m}_{\to \widehat{I}(x)}(x)) & \text{otherwise.} \end{array} \right. \tag{7}$$

Next, if $\widehat{I}(x) + 1 < T$, we have, by construction,

$$\underline{H}_{\widehat{I}(x)+1}(\widehat{m}_{\to \widehat{I}(x)+1}(x)) \times \overline{H}_{\widehat{I}(x)+1}(\widehat{m}_{\to \widehat{I}(x)+1}(x)) = 0.$$

Then, we construct $\widehat{m}_{\widehat{I}(x)+2}, ..., \widehat{m}_T$ by induction using similar formulas as (7).

4. Estimation of the sharp bounds:

   Let us define

$$\widehat{Z}_{it} = \binom{T-t}{S_i - t} \frac{\exp(S_i v(X_i, \widehat{\beta}))}{C_{S_i}(X_i; \widehat{\beta})}.$$

   We estimate $\underline{\text{AME}}_t$ by plug-in: $\widehat{\underline{\text{AME}}}_t = \frac{1}{n}\sum_{i=1}^n \widehat{\underline{h}}(X_i, S_i)$, with

$$\widehat{\underline{h}}(X_i, S_i) = \sum_{t=0}^{T} \lambda_t(X_i, \widehat{\beta}) \widehat{Z}_{it} + \lambda_{T+1}(X_i, \widehat{\beta}) \widehat{Z}_{i0} \times \left[\mathbb{1}\left\{ \lambda_{T+1}(X_i, \widehat{\beta} > 0 \right\} \underline{q}_T(\widehat{m}(X_i)) \right. \tag{8}$$

$$\left. + \mathbb{1}\left\{ \lambda_{T+1}(X_i, \widehat{\beta} < 0 \right\} \overline{q}_T(\widehat{m}(X_i)) \right]. \tag{9}$$

   We define $\widehat{\overline{\text{AME}}}_t$, $\widehat{\underline{\text{ATE}}}_t$ and $\widehat{\overline{\text{ATE}}}_t$ similarly.

Inference on the AME and ATE based on the sharp bounds

Under the conditions of Theorem 3 in DDL, the estimated bounds are asymptotically normal if $\beta_{0k} \neq 0$, but not asymptotically normal in general otherwise. Nevertheless, DDL construct confidence intervals (CI) on AME_t and ATE_t that are asymptotically valid (in a pointwise sense) whether or not $\beta_{0k} = 0$. To present them, let us first define the estimator of the asymptotic variances of the bounds, in case of asymptotic normality. First, let

$$\widehat{\underline{\psi}}_i := \widehat{\underline{h}}(X_i, S_i) - \widehat{\underline{\text{AME}}}_t + \widehat{\underline{v}}'_\beta \widehat{\phi}_i + \widehat{\underline{v}}'_\gamma [\Gamma_i - \widehat{\gamma}(X_i)], \tag{10}$$

where $\Gamma_i = (\mathbb{1}\{S_i = 0\}, ..., \mathbb{1}\{S_i = T\})'$ and

$$\widehat{\phi}_i := -\left[\frac{1}{n} \sum_{j=1}^n \partial^2 \ell_c / \partial\beta\partial\beta'(Y_j | X_j; \widehat{\beta}) \right]^{-1} \partial \ell_c / \partial\beta(Y_i | X_i; \widehat{\beta}).$$

The other terms $\widehat{\underline{v}}_\beta$ and $\widehat{\underline{v}}_\gamma$ are defined in Section D.2 of the Online Appendix of DDL. $\widehat{\overline{\psi}}_i$ is defined similarly. Finally, let

$$\widehat{\underline{\sigma}}^2 := \frac{1}{n} \sum_{i=1}^n \widehat{\underline{\psi}}_i^2, \quad \widehat{\overline{\sigma}}^2 := \frac{1}{n} \sum_{i=1}^n \widehat{\overline{\psi}}_i^2.$$

Note that the first term $\underline{\widehat{h}}(X_i, S_i) - \underline{\widehat{\text{AME}_t}}$ in $\underline{\widehat{\psi}}_i$ would be the influence function of i in the estimation of AME_t if β_0 and $m(\cdot)$ were known. The term $\underline{\widehat{v}}_\beta' \widehat{\phi}_i$ accounts for the influence of i when estimating β_0. The term $\underline{\widehat{v}}_\beta' \widehat{\phi}_i$ accounts for the influence of i in the estimation of β_0. Finally, the term $\underline{\widehat{v}}_\gamma' [\Gamma_i - \widehat{\gamma}(X_i)]$ accounts for the influence of i in the estimation of γ_0.

Now, let φ_α denote a consistent test with asymptotic level α of $\beta_{0k} = 0$, e.g., a t-test. Following Imbens and Manski (2004), let c_α denote the unique solution to

$$\Phi\left(c_\alpha + \frac{n^{1/2}\left(\overline{\widehat{\text{AME}}}_t - \underline{\widehat{\text{AME}}}_t\right)}{\max\left(\underline{\widehat{\sigma}}, \overline{\widehat{\sigma}}\right)}\right) - \Phi(-c_\alpha) = 1 - \alpha, \tag{11}$$

with Φ the cumulative distribution function of a standard normal distribution. Then, define $\text{CI}^1_{1-\alpha}$ as

$$\text{CI}^1_{1-\alpha} := \left| \begin{array}{ll} \left[\underline{\widehat{\text{AME}}}_t - c_\alpha \underline{\widehat{\sigma}}^2/n^{1/2},\ \overline{\widehat{\text{AME}}}_t + c_\alpha \overline{\widehat{\sigma}}/n^{1/2}\right] & \text{if } \varphi_\alpha = 1, \\ \left[\min\left(0, \underline{\widehat{\text{AME}}}_t - c_\alpha \underline{\widehat{\sigma}}/n^{1/2}\right),\ \max\left(0, \overline{\widehat{\text{AME}}}_t + c_\alpha \overline{\widehat{\sigma}}/n^{1/2}\right)\right] & \text{if } \varphi_\alpha = 0. \end{array} \right.$$

2.2. Outer bounds: identification, estimation and inference

Identification

Sharp bounds do not require any optimization but they still involve nonparametri estimation of $m(X)$. We now present simpler outer bounds, which do not require any nonparametric estimation. To keep the exposition concise, we focus on AME_t hereafter. Recall that AME_t is not identified because of the unknown moment $\int_0^1 u^{T+1} d\mu_X(u)$ for some probability measure μ_X on $[0, 1]$ satisfying $\mu_x(\{0, 1\}) = 0$ and $\int_0^1 u^k \mu_x(u) du = m_t(x)$. The idea used by DDL to obtain outer bounds is that if $|u^{T+1} - \sum_{t=0}^{T+1} b_t u^t| \le \eta$ for all u on $[0, 1]$ and some $\eta > 0$, then, by Jensen's inequality,

$$\left|\int_0^1 u^{T+1} d\mu_X(u) - \sum_{t=0}^{T} b_t m_t(X)\right| \le \eta.$$

Also, by definition of $m_t(x)$, we have $\mathsf{E}[Z_0 m_t(X)] = \mathsf{E}[Z_t]$. Then, using (2), we get

$$\left|\text{AME}_t - \mathsf{E}\left[\sum_{t=0}^{T} (\lambda_t(X, \beta_0) + b_t \lambda_{T+1}(X, \beta_0)) Z_t\right]\right| \le \eta \mathsf{E}\left[\,|\lambda_{T+1}(X, \beta_0)| Z_0\right].$$

We can then find the $(b_0, ..., b_T)$ that minimize η. The solution is related to Chebyshev polynomials. Let $\mathbb{T}^c_{T+1}$ denote the Chebyshev polynomial of degree $T+1$, renormalized so that its leading coefficient is equal to 1 (recall that the unnormalized Chebyshev polynomials are defined by $\mathbb{T}^u_0(x) = 1$, $\mathbb{T}^u_1(x) = x$ and $\mathbb{T}^u_{k+1}(x) = 2x\mathbb{T}^u_k(x) - \mathbb{T}^u_{k-1}(x)$ for any $k \ge 1$). Then, let $\mathbb{T}_{T+1}(u) := 2^{-T-1}\mathbb{T}^c_{T+1}(2u - 1)$ and let $-b^*_{k,T}$ denote the coefficient of degree k of $\mathbb{T}_{T+1}$. The η associated to these coefficients is $1/[2 \times 4^T]$. We thus obtain the outer bounds

$\underline{\mathrm{AME}}_t^o := \mathrm{AME}_t^a - \overline{b}$ and $\overline{\mathrm{AME}}_t^o := \mathrm{AME}_t^a + \overline{b}$, with

$$\begin{aligned}\mathrm{AME}_t^a &= \mathsf{E}\left[\sum_{t=0}^{T}(\lambda_t(X,\beta_0) + b_{t,T}^*\lambda_{T+1}(X,\beta_0))Z_t\right],\\ \overline{b} &= \frac{1}{2\times 4^T}\mathsf{E}\left[\,|\lambda_{T+1}(X,\beta_0)|Z_0\right].\end{aligned}$$

Our notation reflects that AME_t^a may be seen as an approximation of the true AME_t, whereas $\overline{b}$ corresponds to the maximal bias of AME_t^a.

Estimation and inference

Let us define

$$p(x,s,\beta) := \sum_{t=0}^{T}(\lambda_t(X,\beta_0) + b_{t,T}^*\lambda_{T+1}(X,\beta_0)\binom{T-t}{s-t}\frac{\exp(sv(x,\beta))}{C_s(x,\beta)},$$

so that $\mathrm{AME}_t^a = \mathsf{E}[p(X_i,S_i,\beta_0]$. We consider plug-in estimators of AME_t^a and $\overline{b}$:

$$\begin{aligned}\widehat{\mathrm{AME}}_t^a &= \frac{1}{n}\sum_{i=1}^{n}p(X_i,S_i,\widehat{\beta}),\\ \widehat{\overline{b}} &= \frac{1}{2\times 4^T n}\sum_{i=1}^{n}\left|\lambda_{T+1}(X_i,\widehat{\beta})\right|\widehat{Z}_{i0}.\end{aligned}$$

Estimated outer bounds are then $(\widehat{\mathrm{AME}}_t^a - \widehat{\overline{b}}, \widehat{\mathrm{AME}}_t^a + \widehat{\overline{b}})$. Then, DDL suggest the following CI on AME_t associated with these outer bounds:

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

where $q_\alpha(b)$ is the quantile of order $1-\alpha$ of a $|\mathcal{N}(b,1)|$ and $\widehat{\sigma}^2$ is an estimator of the asymptotic variance of $\widehat{\mathrm{AME}}_t^a$, $\widehat{\sigma}^2 := \sum_{i=1}^n\widehat{\psi}_i^2/n$ with

$$\widehat{\psi}_i = p(X_i,S_i,\widehat{\beta}) - \frac{1}{n}\sum_{i=1}^{n}p(X_i,S_i,\widehat{\beta}) + \left(\frac{1}{n}\sum_{i=1}^{n}\frac{\partial p}{\partial\beta}(X_i,S_i,\widehat{\beta})\right)'\widehat{\phi}_i.$$

DDL consider another confidence interval, larger than $\mathrm{CI}_{1-\alpha}^2$ but with uniform guarantees:

$$\mathrm{CI}_{1-\alpha}^3 = \left[\widehat{\mathrm{AME}}_t^a \pm \frac{\widehat{\sigma}}{n^{1/2}}q_{\alpha-\alpha_1}\left(\frac{n^{1/2}\widehat{\overline{b}}_{\alpha_1}}{\widehat{\sigma}}\right)\right],$$

for some $\alpha_1\in(0,\alpha)$. The construction of $\widehat{\overline{b}}_{\alpha_1}$, which is the upper bound of a CI on $\overline{b}$ with nominal coverage $1-\alpha_1$, is detailed in Appendix B of DDL.

2.3. A note on clustering

We have assumed thus far to have an i.i.d. sample. Correlation between individuals does not affect identification of the sharp and outer bounds, but it affects the asymptotic variance of the estimators and hence the confidence intervals presented before. With clustering, the formulas remain the same except that, to account for clustering, individual influence functions are summed up within each cluster, before the variance of these clustered influence functions are computed.

3. The functions in the MarginalFElogit package

We now present the core function of the package, namely `felogit`, which estimates the AME or ATE for the specified covariates at given periods. Note that the `summary_felogit` function displays the result from the `felogit` function in a `knitr` table. After having described `felogit`, we discuss the preprocessing performed on the data before estimation. We conclude this section with a number of practical recommendations regarding the usage of the package.

3.1. The felogit function

We detail below the arguments and outputs of the `felogit` function.

Arguments

The function takes the following parameters as inputs:

`data` a data frame containing the panel data to perform estimation upon. The data can be in long form, with each row giving the values of the variables observed at a given period for a given individual. In this case, the first column should record the unique identifier of the individual observed and the second column should record the period of observation. All variables should be named. The variables considered are given in the `formul` parameter described below. Alternatively, `data` may already be in wide format, a list containing the matrix Y as its first element (the element should be named so that we have a name for the variable) and the array X as its second element, with names along the third dimension so that we may know what name to give to the variables. In both cases, rows should represent individuals and columns should represent periods. The values recorded in Y should represent the (binary) variable of interest while each slice of X along the third dimension should record the values of one explanatory variable. `NA`s are accepted values and represent missing observations.

`formul` *(optional, default* `NULL`*)* an object of class "formula". If `response` is the name of the (binary) variable of interest in `data`, and `terms` are the names of the explanatory used to fit the logit model, one can specify it as e.g. `response` $\sim$ `terms`. Other formulae are acceptable as long as the first variable that appears is the variable of interest, and the others are the ones to use to fit the model. Note that constant-through-time covariates will be discarded. `formul` can also be set to `NULL` to imply `data` is already in wide format, as explained above.

`Option` *(optional, default "*`quick`*")* the method used to fit the AME/ATE. If "`quick`" or "`outer`" (case-insensitive) the outer bounds methods from 2.2 is used, otherwise (e.g. "`sharp`"), the sharp bounds method from 2.1 is used.

`compute_X` *(optional, default* `NULL`*)* the vector containing the names of the explanatory variables with respect to which to compute the bounds on the AME or ATE. If `NULL` the bounds are computed with respect to all covariates.

`compute_T` *(optional, default* `NULL`*)* the vector of periods at which to compute the AMEs or ATEs. Individuals unobserved at one of the period will be dropped to estimate the relevant AME. Zero or negative values may be specified, in which case if the value is $-k$ the AMEs/ATEs will be computed using, for each individual, the period k periods earlier than its last period of observation, in an event-study fashion. If specified to "`all`", the AMEs/ATEs are computed at each fixed period, and the average effect is also computed. Here, by average effect, we mean the average across individuals of the individual time-averages, using for each individual all periods at which it can be observed. If `NULL`, we use the earliest period for which observations are available.

`cluster` *(optional, default* `NULL`*)* the vector of identifiers for the clusters. Its length should be the number of individuals and cluster identifiers are listed in increasing order of the individual identifiers. Individuals sharing the same identifier belong to the same cluster and may present correlated covariates or individual fixed effects. If `NULL`, each individual is its own cluster, i.e. individual observations are independent.

`alpha` *(optional, default* `5`*)* one minus the nominal coverage for the CI. Default is 5% (thus corresponding to a nominal coverage of 95%).

`CIOption` *(optional, default "*`CI2`*")* the confidence interval to be used for the outer bounds method. Default ("CI2") computes the $\mathrm{CI}^2_{1-\alpha}$ confidence interval, any other value will compute $\mathrm{CI}^3_{1-\alpha}$ instead. This parameter is ignored if `Option` is "`sharp`" (or any value other than "`quick`" or "`outer`" up to changes in the case).

`nbCores` *(optional, default 4)* the number of cores used for parallel computing in the sharp bounds methods. Parallel computing is used to speed up the estimation of the sharp bounds, this parameter is ignored if `Option` is "`quick`" or "`outer`", up to changes in the case.

Outputs

The function outputs a list containing the following variables:

`summary` a matrix containing the results from our partial identification strategy. For each period in `compute_T` and each variable in `compute_X`, the matrix shows whether the estimated bounds correspond to an AME or to an ATE (i.e., whether the variable is continuous or binary), the estimated bounds on that parameter, and the requested confidence interval. If `compute_T` was set to "`all`", the average (in the sense described in the description of the `compute_T` argument) for each explanatory variable in `compute_X` is also displayed.

`nobs` the number of individuals used for estimation (individuals observed for at least two periods).

`ndiscard` the number of individuals discarded from the data (individuals observed for only one period, or none).

`Tmax` the number of distinct periods observed in the data.

`vardiscard` a vector containing the names of the variables, among those given in `formul`, that were dismissed because, for more than 99% of the population, they displayed no time variation for each given individual.

`formul` the formula `formul` given in input.

`alpha` level of the confidence intervals. Same as input, or 5% if input `alpha` was `NULL`.

`summary_CMLE` a matrix containing the estimates from the CMLE estimation. For each explanatory variable k, the matrix contains the point estimate for β_{0k}, the estimated variance on that estimate, and the corresponding p-value for a Wald test against $\beta_{0k} = 0$.

`compute_T` periods at which the AMEs/ATEs were computed. Same as input.

The list returned by the `felogit` function can be passed on to the `summary_felogit` for the results to be nicely displayed. One can also add the argument `format = "latex"` to obtain the tables in LaTeX format.

3.2. Data preprocessing

The data is preprocessed in the following order.

1. If `data` is provided in long format, the input parameter `formul` needs to be specified. We extract the variables whose name appears in `formul` to convert `data` into wide formats, dropping out the variables not appearing in `formul`.
2. Individual-period pairs with at least one variable missing are considered fully unobserved and discarded.
3. Variables which do not vary across periods for over 99% of individuals (ignoring missing observations) are discarded.
4. Individuals which are observed for only one period are discarded.
5. Infer the type of each variable from the observed values: variables for which only two values are observed are considered binary, the others are deemed continuous. If any of the variables for which the effect must be computed is binary, we reverse back to the outer bounds estimation method, independently from the user query.

3.3. Practical recommendations relative to efficiency and robustness

Choosing between the two methods. the estimates of the outer bounds and the corresponding CIs are much quicker to compute than the sharp bounds and $\mathrm{CI}^1_{1-\alpha}$. Moreover, the simulations in DDL suggest that the loss in terms of length of the corresponding CIs can be expected to be modest in most cases. The only case where we recommend using the sharp bounds is with only one continuous regressor, few discrete regressors, n large (greater than 5,000) and $T = 2$.

Attrition and missing data In their Online Appendix A.3, DDL point out that their methodology remains valid even when the number of observed periods varies across individuals, by just applying the same formulas using the value of T corresponding to each individual. In particular, attrition may be correlated with the covariates and the individual effects without threatening the validity of the method. Our implementation of the code supports any attrition pattern, meaning we also accept individuals that are unobserved at a period in-between two periods at which they are observed. However, for the nonparametric estimation used to estimate the sharp bounds, we do not run local linear regressions only between individuals with the same attrition pattern. This is because some attrition patterns may only apply to a few individuals, and then the predicted value would be unstable. Nonetheless, in doing so, we implicitly assume that individuals with different attrition patterns are comparable. Therefore, in our implementation of the sharp bounds method and unlike DDL, we assume that attrition is independent from the observed covariates and the individual fixed effects. Outer bounds do not rely on this independence assumption, however.

Note also that individual-period pairs with one missing covariate are considered unobserved as a whole. Thus, if many covariates with missing values are used, one may end up excludig many individual-period cells from the estimation.

4. Examples

4.1. Simulated dataset

We consider similar simulations as DDL. The corresponding data are generated through the function `generate_data()`. Below we consider DGP2 of DDL with $n = 500$ and $T = 3$. This means that $\beta_0 = 1$, (X_1, X_2, X_3) are i.i.d. with $X_t \in \mathbb{R}$, uniformly distributed on $[-1/2, 1/2]$ and $\alpha = -X_T + \eta$ with $\eta|X \sim \mathcal{N}(0, 1)$. In this case, the AME is equal to 0.2066 at the three periods. The sharp bounds are [0.2059, 0.2069] and the outer bounds are [0.2058, 0.2076].

```
rm(list = ls())
gc()

# Load packages
library(rlang)
library(pracma)
library(snowfall)
library(kableExtra)
library(tidyverse)
library(R.matlab)
invisible(lapply(list.files(paste0(getwd(), "/R/code_restructure"),
                                pattern = ".R$", full.names = TRUE), source))
invisible(source("R/code_restructure/simulations/generate_data.R"))

# Generate data
data <- generate_data(500, 3, DGP = 2, beta = 1)\$data

# Perform estimation and display results
```

```
output <- felogit(data)
summary_felogit(output)
```

The function `felogit` returns a vector containing a summary table of the results ("summary"), the number of used individuals ("n"), the number of discarded individuals ("ndiscard"), the maximal number of periods of observed ("Tmax"), the label of the discarded variables ("vardiscard"), the formula used ("formul"), the level used for the confidence intervals ("alpha") as well as the method ("Option"). The function `summary_felogit` applied to the output of `felogit` displays in the console the following table:

```
Estimates of coefficients in the fixed effect logit model (CMLE)

========  ==========  ==============  ======
Variable  Point Est.  se(Point Est.)  pvalue
========  ==========  ==============  ======
X_1       1.2305      0.2399          0
========  ==========  ==============  ======
Estimates of the Average Marginal/Treatment Effects in the fixed effect logit
model

=======  ========  =======  ========================  ================
Period   Variable  AME/ATE  Estimated (outer) bounds  95% CI
=======  ========  =======  ========================  ================
T=1      X_1       AME      [0.2501, 0.2539]          [0.1608, 0.3431]
T=2      X_1       AME      [0.2516, 0.2555]          [0.161, 0.3461]
T=3      X_1       AME      [0.2685, 0.2721]          [0.1662, 0.3745]
Average  X_1       AME      [0.2567, 0.2605]          [0.163, 0.3542]
=======  ========  =======  ========================  ================
Formula: Y ~ X_1.
Number of discarded individuals: 0.
Number of non-discarded individuals: 500.
Number of observed periods: 3.
The method used to compute the AME/ATE is the "quick" method (i.e. the second
method in DDL).
The column period indicates at which period the AME/ATE are computed.
Average corresponds to the average of AME/ATE over the different periods.
CI: Confidence intervals are computed at a 5% level.
```

Note that other formats are available, either in the viewer using `summary_felogit(output, format="html")` or in latex using `summary_felogit(output, format="latex")`.

4.2. Real dataset

For illustration purposes, we consider the real dataset UnionWage from the package pglm, which is consists in 545 yearly individuals observations from 1980 to 1987 of Unionism in the United States (see ?UnionWage for more details). Consider a panel logit model for whether the wage is set by collective bargaining or not (binary variable "union") and as predictors the marital status (variable “married"), the experience, computed as age - 6 - schooling (variable “exper"). Below, for illustration purposes, we also include the binary variable "black" indicating the black community, which will be discarded because it is constant over time.

```
rm(list = ls())
gc()

# Load packages
library(rlang)
library(pracma)
library(snowfall)
library(kableExtra)
library(tidyverse)
library(R.matlab)
invisible(lapply(list.files(paste0(getwd(), "/R/code_restructure"),
                        pattern = ".R$", full.names = TRUE), source))

# Load data
data('UnionWage', package = 'pglm')

# Subset the dataframe
subData <- UnionWage[UnionWage\$year < 1986, c("id", "year",
        "exper", "married", "union", "wage", "com", "region")]
subData\$union <- subData\$union == "yes"
subData\$black <- subData\$com == "black"
subData$NorthEast <- subData$region == "NorthEast"
formul <- formula(union ~ exper + married + black)

# Compute outer bounds at all periods, plus the average of all periods
output <- felogit(data = subData, formul = formul, Option = "quick",
                                compute_T = "all")
summary_felogit(output)
```

The table produced by `summary_felogit()` in the R viewer is displayed below. Figure 1 also shows the output in html format.

```
Estimates of coefficients in the fixed effect logit model (CMLE)

========  ==========  ==============  ======
Variable  Point Est.  se(Point Est.)  pvalue
========  ==========  ==============  ======
exper     -0.0612     0.0325          0.0596
married   0.1599      0.2041          0.4334
========  ==========  ==============  ======
Estimates of the Average Marginal/Treatment Effects in the fixed effect
logit model
```

Period	Variable	AME/ATE	Estimated (outer) bounds	95% CI
T=1980	exper	AME	[-0.0053, -0.0053]	[-0.0083, -0.0022]
	married	ATE	[0.019, 0.019]	[-0.0739, 0.1119]
T=1981	exper	AME	[-0.0052, -0.0052]	[-0.0081, -0.0023]
	married	ATE	[-0.0039, -0.0039]	[-0.0602, 0.0524]
T=1982	exper	AME	[-0.0051, -0.0051]	[-0.008, -0.0023]
	married	ATE	[0.0238, 0.0238]	[-0.0265, 0.0742]
T=1983	exper	AME	[-0.0051, -0.0051]	[-0.008, -0.0023]
	married	ATE	[-0.0097, -0.0097]	[-0.0483, 0.0288]
T=1984	exper	AME	[-0.005, -0.005]	[-0.0078, -0.0023]
	married	ATE	[0.0211, 0.0211]	[-0.0168, 0.059]
T=1985	exper	AME	[-0.005, -0.005]	[-0.0077, -0.0023]
	married	ATE	[0.0308, 0.0308]	[-0.0042, 0.0659]
Average	exper	AME	[-0.0051, -0.0051]	[-0.008, -0.0023]
	married	ATE	[0.0135, 0.0135]	[-0.0221, 0.0491]

```
Formula: union ~ exper + married.
Number of discarded individuals: 0.
Number of non-discarded individuals: 545.
Number of observed periods: 6.
The method used to compute the AME/ATE is the "quick" method (i.e. the
second method in DDL).
The column period indicates at which period the AME/ATE are computed.
Average corresponds to the average of AME/ATE over the different periods.
CI: Confidence intervals are computed at a 5% level.
The variable black has been discarded because it is constant or almost
constant over time.
```

Estimates of the Average Marginal Effects in the fixed effect logit model

Variable	Point Est.	se(Point Est.)	pvalue
exper	-0.0612	0.0325	0.0596
married	0.1599	0.2041	0.4334

Period	Variable	AME/ATE	Estimated (outer) bounds	95% CI
T=1980	exper	AME	[-0.0053, -0.0053]	[-0.0083, -0.0022]
	married	ATE	[0.019, 0.019]	[-0.0739, 0.1119]
T=1981	exper	AME	[-0.0052, -0.0052]	[-0.0081, -0.0023]
	married	ATE	[-0.0039, -0.0039]	[-0.0602, 0.0524]
T=1982	exper	AME	[-0.0051, -0.0051]	[-0.008, -0.0023]
	married	ATE	[0.0238, 0.0238]	[-0.0265, 0.0742]
T=1983	exper	AME	[-0.0051, -0.0051]	[-0.008, -0.0023]
	married	ATE	[-0.0097, -0.0097]	[-0.0483, 0.0288]
T=1984	exper	AME	[-0.005, -0.005]	[-0.0078, -0.0023]
	married	ATE	[0.0211, 0.0211]	[-0.0168, 0.059]
T=1985	exper	AME	[-0.005, -0.005]	[-0.0077, -0.0023]
	married	ATE	[0.0308, 0.0308]	[-0.0042, 0.0659]
Average	exper	AME	[-0.0051, -0.0051]	[-0.008, -0.0023]
	married	ATE	[0.0135, 0.0135]	[-0.0221, 0.0491]

Note:
Left table: Estimates of coefficients in the fixed effect logit model (CMLE)
Right table: Estimates of the Average Marginal Effects
Formula: union ~ exper + married.
Number of discarded individuals: 0.
Number of non-discarded individuals: 545.
Number of observed periods: 6.
The method used to compute the AME/ATE is the "quick" method (i.e. the second method in DDL).
The column period indicates at which period the AME/ATE are computed.
Average corresponds to the average of AME/ATE over the different periods.
CI: Confidence intervals are computed at a 5% level.
The variable black has been discarded because it is constant or almost constant over time.

Figure 1: Output of `summary_felogit()` on the real UnionWage dataset.

For comparison purposes, the output of the conditional logit estimator of $\widehat{\beta}$ can be compared to the one of the `clogit()` function in the package survival, using the following syntax:

```
fe <- clogit(union ~ exper + married + strata(id), data = sub)
summary(fe)
```

References

Andersen EB (1970). "Asymptotic Properties of Conditional Maximum-Likelihood Estimators." *Journal of the Royal Statistical Society. Series B (Methodological)*, **32**(2), 283–301.

Davezies L, D'Haultfœuille X, Laage L (2024). "Identification and estimation of average Causal effects in fixed effect logit models." ArXiv preprint arXiv:2105.00879.

Imbens GW, Manski CF (2004). "Confidence intervals for partially identified parameters." *Econometrica*, **72**(6), 1845–1857.

Schmüdgen K (2017). *The moment problem*, volume 14. Springer.

Affiliation:

Laurent Davezies
CREST-ENSAE
5, avenue Henry Le Chatelier
91 120 Palaiseau, France
E-mail: laurent.davezies@ensae.fr

Xavier D'Haultfœuille
CREST-ENSAE
5, avenue Henry Le Chatelier
91 120 Palaiseau, France
E-mail: xavier.dhaultfoeuille@ensae.fr

Christophe Gaillac
University of Geneva
Uni Mail - 40 bd du Pont-d'Arve
CH-1211 Genève 4, Switzerland
E-mail: christophe.gaillac@unige.ch

Louise Laage
Georgetown University
E-mail: louise.laage@georgetown.edu

Maël Laoufi
CREST-ENSAE
5, avenue Henry Le Chatelier
91 120 Palaiseau, France
E-mail: mael.laoufi@ensae.fr