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

**Laurent Davezies**
CREST

**Xavier D'Haultfœuille**
CREST

**Christophe Gaillac**
CREST

**Louise Laage**
Georgetown University

### Abstract

This vignette presents the package **MarginalFElogit** associated to Davezies, D'Haultfoeuille, and Laage (2021) (DDL hereafter). It considers inference about average marginal effects (AME) and similar parameters in a panel data fixed effects logit model. It implements the nonparametric estimator of the sharp bounds on the AME and the related confidence intervals on the AME from DDL. It also implements the second method proposed in DDL, which does not require any nonparametric estimation but may result in larger confidence intervals. This paper illustrates the usage of **MarginalFElogit** with several simulated and real examples. R and the package **MarginalFElogit** are open-source software projects and can be freely downloaded respectively from CRAN: http://cran.r-project.org and the Github https://github.com/cgaillac/MarginalFElogit.

*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. 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 and observe binary outcomes $Y_1, ..., Y_T$ and for each period $t$, a vector of covariates $X_t := (X_{t1}, ..., X_{tp})'$. Let $Y := (Y_1, ..., Y_T)'$, $X := (X_1', ..., X_T')'$ and consider the logit model with fixed effects:

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

Importantly, the individual effect $\alpha$ is allowed to be correlated in an unspecified way with $X$. DDL provide methods to compute the bounds on two usual parameters of interest:

1. the Average Marginal Effect (AME) $\Delta$, *i.e.*, 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

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

2. the Average Treatment Effect (ATE) $\Delta^{ATE}$, *i.e.*, the effect on $Y_T$ of a universal, exogenous change in $X_{kT}$ from 0 to 1,

$$
\Delta_k^{ATE} := \mathsf{E}\left[\Lambda(X_{T-k}'\beta_{0-k} + \beta_{0k}) - \Lambda(X_{T-k}'\beta_{0-k})\right].
$$

We present here the two methods of DDL to compute the bounds on the AME. We refer to Section 5.1 in DDL which simply explains the changes to compute the ones on the ATE. Let us define $S = \sum_{t=1}^T Y_t$, $(\lambda_t)_{t=0}^{T+1}$ from

$$
\sum_{t=0}^{T+1} \lambda_t(x, \beta_0)u^t := u(1-u)\prod_{t=1}^{T-1}\left(u(\exp((x_t - x_T)'\beta_0) - 1) + 1\right),
\tag{2}
$$

$$
c_t(x) := \mathsf{E}\left[\frac{\mathbb{1}\{S \geq t\}\binom{T-t}{S-t}\exp(Sx_T'\beta_0)}{C_S(x, \beta_0)}\Big|X = x\right],\ t \in \{0, ..., T\}
\tag{3}
$$

$$
C_k(x, \beta) := \sum_{(d_1, ..., d_T)\in\{0,1\}^T:\sum_{t=1}^T d_t = k} \exp\left(\sum_{t=1}^T d_t x_t'\beta\right),
\tag{4}
$$

$$
m_t(x) := \frac{c_t(x)}{c_0(x)},\ t \in \{0, ..., T\},\ m(x) := (m_0(x), ..., m_T(x))'.
\tag{5}
$$

For any $m \in [0,1]^{T+1}$ we denote by $\mathcal{D}(m)$ the set of positive measures $\mu$ on $[0,1]$ whose vector of first $T+1$ raw moments (starting from $\int_0^1 u^0 d\mu(u) = \mu([0,1])$) is equal to $m$. Finally, define

$$
\underline{q}_T(m) := \inf_{\mu \in \mathcal{D}(m)}\int_0^1 u^{T+1}d\mu(u),\ \overline{q}_T(m) := \sup_{\mu \in \mathcal{D}(m)}\int_0^1 u^{T+1}d\mu(u).
\tag{6}
$$

From Lemma 1 in DDL, the sharp identified set of $\Delta(x)$ is $[\underline{\Delta}(x), \overline{\Delta}(x)]$, with

$$
\begin{cases}
\underline{\Delta}(x) &= \beta_{0k}\left[\sum_{t=1}^T \lambda_t(x, \beta_0)c_t(x) + c_0(x)\lambda_{T+1}(x, \beta_0)\underline{q}_T(m(x))\right] \\
\overline{\Delta}(x) &= \beta_{0k}\left[\sum_{t=1}^T \lambda_t(x, \beta_0)c_t(x) + c_0(x)\lambda_{T+1}(x, \beta_0)\overline{q}_T(m(x))\right]
\end{cases},
\tag{7}
$$

if $\beta_{0k}\lambda_{T+1}(x, \beta_0) \geq 0$. If $\beta_{0k}\lambda_{T+1}(x, \beta_0) < 0$, the same holds with $\overline{q}_T$ and $\underline{q}_T$ switched in the two bounds.

## 2.1. The first "sharp" estimation and inference method

The first methods in DDL estimates the sharp bounds on $\Delta$ and develop inference on this parameter based on these bounds, using a sample $(Y_i, X_i)_{i=1,...,n}$. These bounds are estimated in three steps:

- **Step 1:** Estimation of $\beta_0$ by the conditional maximum likelihood estimator $\widehat{\beta}$.

- **Step 2:** Estimation of the functions $c_0, ..., c_T$ and $m$:

  (a) Nonparametric estimation of $c_0, ..., c_T$;
  (b) Nonparametric estimation of $m$ by an estimator $\widehat{m}$ satisfying, for all $i$, $\widehat{m}(X_i) \in \mathcal{M}_T$, the set of all possible vectors of first $t+1$ raw moments, starting from the moment of order 0, of probability measures on $[0, 1]$.

- **Step 3:** Estimation of the bounds by a plug-in estimator

$$\widehat{\overline{\Delta}} = \frac{1}{n}\sum_{i=1}^{n} U(X_i, S_i, \widehat{\beta}) + \widehat{\beta}_k\widehat{c}_0(X_i)\lambda_{T+1}(X_i, \widehat{\beta})\Big[\overline{q}_T(\widehat{m}(X_i))\mathbb{1}\left\{\widehat{\beta}_k\lambda_{T+1}(X_i, \widehat{\beta}) \geq 0\right\}$$

$$+ \underline{q}_T(\widehat{m}(X_i))\mathbb{1}\left\{\widehat{\beta}_k\lambda_{T+1}(X_i, \widehat{\beta}) < 0\right\}\Big], \tag{8}$$

where

$$U(x, s, \beta) := \beta_k \sum_{t=0}^{s}\binom{T-t}{s-t}\frac{\lambda_t(x; \beta_0)\exp(sx_T'\beta)}{C_s(x, \beta)}$$

and similarly for $\widehat{\underline{\Delta}}$. Step 1 is straightforward. We now explain in details the two parts of Steps 2.

*Estimation of $(c_0, ..., c_T)$*

Let $\gamma_{0j}(x) = P(S = j | X = x)$ for $j = 0, ..., T$. The functions $(c_t)_{t=0...T}$ and $(\gamma_{0j})_{j=0...T}$ are related through

$$(c_0(x), ..., c_T(x))' = \Gamma\left(\frac{\gamma_{00}(x)\exp(0 \times x_T'\beta_0)}{C_0(x, \beta_0)}, \ ... \ , \frac{\gamma_{0T}(x)\exp(T \times x_T'\beta_0)}{C_T(x, \beta_0)}\right)', \tag{9}$$

where $\Gamma$ is a square matrix of size $T+1$ with coefficients $\Gamma_{ij} = \binom{T-i}{j-i}\mathbb{1}\{i \leq j\}$ for $i, j = 1, ..., T+1$. We first estimate $\gamma_0 := (\gamma_{00}, ..., \gamma_{0T})$ nonparametrically. We use local polynomial estimators of order $\ell$ to avoid boundary effects. Let $K$ denote a kernel function and for a given $0 \leq j \leq T$, define

$$\widehat{a}^j(x) := \text{argmin}_a \sum_{i=1}^{n} K\left(\frac{X_i - x}{h_n}\right)\left(\mathbb{1}\{S_i = j\} - \sum_{|b| \leq \ell} a_b(X_i - x)^b\right)^2, \tag{10}$$

where, in this definition, $b \in \mathbb{N}^{pT}$, $|b| = \sum_{j=1}^{pT} b_j$ and $x^b = x_1^{b_1} ... x_{pT}^{b_{pT}}$. The estimator of $\gamma_{0j}(x)$ is then $\widehat{\gamma}_j(x) = \widehat{a}_0^j(x)$. The estimator for $c_t(x)$, $\widehat{c}_t(x)$, uses (9), replacing $\gamma_0$ and $\beta_0$ with their estimators.

*Estimation of m*

Given the definition of $m$, a natural estimator is

$$\widetilde{m}(x) = \left(1, \ \frac{\widehat{c}_1(x)}{\widehat{c}_0(x)}, \ ..., \ \frac{\widehat{c}_T(x)}{\widehat{c}_0(x)}\right).$$

However, this estimator may not satisfy $\widetilde{m}(x) \in \mathcal{M}_T$. We thus consider another estimator $\widehat{m}$ such that $\widehat{m}(x) \in \mathcal{M}_T$. Specifically, let $c_n$ be a sequence tending to 0 at a rate specified later and define

$$\widehat{I}(x) := \max\left\{t \in \{1, ..., T\} : \underline{H}_t(\widetilde{m}_{\rightarrow t}(x)) \times \overline{H}_t(\widetilde{m}_{\rightarrow t}(x)) > c_n\right\}.$$

with the convention that $\max \emptyset = 0$. We then let

$$\widehat{m}_{\rightarrow \widehat{I}(x)}(x) := \widetilde{m}_{\rightarrow \widehat{I}(x)}(x).$$

If $\widehat{I}(x) = T$, $\widehat{m}(x)$ is fully defined. Otherwise, we complete $\widehat{m}(x)$ by first letting

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

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

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

Then, we construct by induction the unique possible moments $\widehat{m}_{\widehat{I}(x)+2}, ..., \widehat{m}_T$ that are compatible with $\widehat{m}_{\rightarrow \widehat{I}(x)+1}(x)$. By construction, the corresponding vector $\widehat{m}(x)$ belongs to $\mathcal{M}_T$.

Under the conditions of Theorem 1 in DDL, the estimated bounds are asymptotically normal if $\beta_{0k} \neq 0$, but not in general if $\beta_{0k} = 0$. An exception is when the whole vector $\beta_0$ is equal to 0. DDL provide confidence intervals on $\Delta$ that are asymptotically valid whether or not $\beta_{0k} = 0$, at least in a pointwise sense. To this end, let $\varphi_\alpha$ denote a consistent test with asymptotic level $\alpha$ of $\beta_{0k} = 0$, *e.g.*, a $t$-test. Following Imbens and Manski (2004), let $c_\alpha$ denote the unique solution to

$$\Phi\left(c_\alpha + \frac{n^{1/2}\left(\widehat{\overline{\Delta}} - \widehat{\underline{\Delta}}\right)}{\max\left(\widehat{\Sigma}_{11}^{1/2}, \widehat{\Sigma}_{22}^{1/2}\right)}\right) - \Phi(-c_\alpha) = 1 - \alpha,$$

with $\Phi$ the cdf of a standard normal distribution and $\Sigma_{ij}$ is the $(i, j)$ term of $\Sigma$. Then, define $\text{CI}_{1-\alpha}^1$ as

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

## 2.2. The "quick" alternative estimator and inference method

One insight from DDL is that the lack of identification of $\Delta$ arises because of the unknown moment $\int_0^1 u^{T+1} d\mu(u)$ for some $\mu \in \mathcal{D}(m)$ made precise in DDL. The idea of the alternative estimator is to use a (good) approximation of $u \mapsto u^{T+1}$ by a polynomial of degree $T$:

$$b^* = \operatorname{argmin}_{b \in \mathbb{R}^{T+1}} \sup_{u \in [0,1]} \left| u^{T+1} - \sum_{k=0}^{T} b_k u^k \right| \tag{11}$$

to obtain, using

$$a_t(x) = \lambda_t(X) + b_t^* \lambda_{T+1}(X), \ t \in \{0, ..., T\},$$

the following approximation of $\Delta$,

$$
\begin{aligned}
\tilde{\Delta} &= \beta_{0k} E\left[ \sum_{t=0}^{T} a_t(X, \beta_0) c_t(X) \right] \\
&= \beta_{0k} E\left[ \sum_{t=0}^{S} \frac{\exp(S X_T' \beta_0) a_t(X, \beta_0) \binom{T-t}{S-t}}{C_S(X, \beta_0)} \right].
\end{aligned}
$$

DDL then use a plug-in estimator of $\tilde{\Delta}$:

$$\widehat{\Delta} = \frac{\widehat{\beta}_k}{n} \sum_{i=1}^{n} \sum_{t=0}^{S_i} \frac{\exp(S_i X_{iT}' \widehat{\beta}) a_t(X_i, \widehat{\beta}) \binom{T-t}{S_i-t}}{C_{S_i}(X_i, \widehat{\beta})}. \tag{12}$$

The approximation $\tilde{\Delta}$ is biased for $\Delta$, but DDL provide a bound $\bar{b}$ on $|\tilde{\Delta} - \Delta|$ which can be estimated consistently by:

$$\widehat{\bar{b}} = \frac{|\widehat{\beta}_k|}{2 \times 4^T} \frac{1}{n} \sum_{i=1}^{n} |\widehat{\lambda}_{T+1}(X_i)| \binom{T}{S_i} \exp(S_i X_{iT}' \widehat{\beta}).$$

Provided that $|\tilde{\Delta} - \Delta| < \bar{b}$ or $\beta_{0k} = 0$, the confidence interval $\text{CI}_{1-\alpha}^2$ defined by

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

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 variance of the influence function of $\widehat{\Delta}$ (provided in Section 4.2 n DDL) has level greater than $1 - \alpha$.

DDL also consider

$$\text{CI}_{1-\alpha}^3 = \left[ \widehat{\Delta} \pm q_\alpha \left( \frac{n^{1/2} \widehat{\bar{b}} + \varepsilon_n}{\widehat{\sigma}} \right) \frac{\widehat{\sigma}}{n^{1/2}} \right],$$

for some $\varepsilon_n \to \infty$ (possibly very slowly, e.g. $\varepsilon_n = \ln \ln(n)$), which has level greater than $1 - \alpha$ even when $\left| \Delta - \tilde{\Delta} \right| = \bar{b}$.

## 3. The functions in the MarginalFElogit package

### 3.1. The felogit function

This function implements the estimators proposed in DDL. The syntax of the function `felogit()` is as follows:

`felogit(data, formul, Option, compute_X, compute_T, cluster, alpha, CIOption, nbCores)`

| | |
|---|---|
| data | a data frame containing the panel data in long, *i.e.* one line for one individual-time observation. The first column must be the individual identifier and the second column the temporal one. It can contains NA, which are treated as explained below. |
| formul | an object of class "formula": a symbolic description of the model to be fitted. Models for `felogit()` are specified symbolically. A typical model has the form "response ~ terms" where response is the (numeric) response vector and terms is a series of terms which specifies a linear predictor for response. Note that predictors which are constant over time are discarded. |
| Option | the method chosen to compute the AME/ATE, either "sharp" (see Section 2.1) or "quick" (see Section 2.2). Default is "quick". |
| compute_X | the vector of selected covariates to compute the AME/ATE. If NULL then bounds are computed for all covariates. NULL by default. |
| compute_T | the vector of selected periods to compute the AME/ATE. If NULL, then as described in Section 5.4 of DDL, AME/ATE are computed at $\underline{T} = \min \operatorname{supp}(T)$. If specified to "all", AME/ATE are computed at all available periods but the average over the latter is also computed. Default is NULL. |
| cluster | the vector of identifiers for the clusters if any. Default is NULL. |
| alpha | the confidence level for the confidence intervals. Default is 5%. |
| CIOption | the option for the choice of the type of confidence intervals for the "quick" method, either $\mathrm{CI}_{1-\alpha}^{2}$ ("CI2") or $\mathrm{CI}_{1-\alpha}^{3}$ ("CI3"). Default is "CI2". |
| nbCores | the number of cores used by the program to compute the AME for the "sharp" method. To reduce the computational time, this function can use several cores, in which case the library `snowfall` should be loaded first. By default, `nbCores` is set to 4. |

We also refer to the reference manuel or help file for additional details.

The `felogit()` function starts by discarding predictors which are constant over time (the chosen criterion is that the 99% quantile of the standard deviation of the associated marginal should be different from zero). Then, it reshapes the data in wide, filling the unobserved period with NAs. If any, AME/ATE are then computed conditionning on the set of observed periods. The number of discarded/observed individals is returned. The `felogit()` function then identifies the type of the predictors, either continuous or binary, and compute the AME/ATE at the selected periods for the selected covariates.

The `felogit()` function returns a vector containing a dataframe `summary` which can be represented as a `knitr` table using the `summary_felogit()`. There, the table is represented in the viewer.

## 4. Examples

## 4.1. On a simulated dataset

We consider the same DGP as in DDL, which are available through the function `generate_data()`. Below we select the DGP 2, *i.e.*, assuming that $(X_1, ..., X_T)$ are i.i.d., with $X_t \in \mathbb{R}$, uniformly distributed on $[-1/2, 1/2]$ and $\beta_0 = 1$. The distribution of $\alpha$ is discrete and takes two values:

$$\alpha = X_T + \eta, \quad P(\eta = -1|X_1, ..., X_T) = P(\eta = 1|X_1, ..., X_T) = 1/2.$$

Because $|\text{Supp}(\alpha|X)| = 2$, $\Delta_1 \simeq 0.1904$ is partially identified if $T < 4$ and point identified otherwise. The true bounds are $[0.1826, 0.1953]$ if $T = 2$ and $[0.1895, 0.1906]$ if $T = 3$.

```
rm(list=ls())
## load packages
library(MarginalFElogit)

library(R.matlab)
library(pracma)
library(MASS)
library(snowfall)
library(survival)

## Data generating process as in DDL
# DGP=1 if alpha=0, 2 if alpha in {-1,0,1}
# and 3 if it is a mixture of continuous and discrete.
DGP=c(2)
n=c(500)

## possible values for Ti and proportions in the population
grid_T=c(3)
prop_T = c(1)

## Number of covariates and true value
dimX=1;
beta0 = c(1)

## continuous 1/binary 0
type_cont = c(1)

## number of cores
nbCores = 5

############# DGP ##############
data <- generate_data(grid_T,prop_T,n,dimX,DGP,beta0,type_cont)

############# Estimation #############
output <- felogit(data,Option ="quick")
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 table

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


======== ========== ============== ======
Variable  Point Est.  se(Point Est.)  pvalue
======== ========== ============== ======
Xc1       0.9128      0.2544          3e-04
======== ========== ============== ======
Estimates of the Average Marginal Effects in the fixed effect logit model


====== ======== ======= =============== ===============
Period  Variable  AME/ATE  Estimate         95% CI
====== ======== ======= =============== ===============
Tinf    Xc1       AME      [0.1831,0.1833]  [0.0832,0.2833]
====== ======== ======= =============== ===============
Formula : Y ~ X1.
Nb of discarded individuals: 0.
Nb of observed individuals: 500.
Maximal number of observed periods: 3.
The method used to compute AME/ATE is the "sharp" method (i.e. the first method
in DDL).
The column period indicates at which period the AME/ATE are computed.
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. On a 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())
## load packages
library(MarginalFElogit)
library(pglm)
```

```
library(plm)
data('UnionWage', package = 'pglm')

## subset of the data frame
sub <- UnionWage[UnionWage\$year <1986 ,c("id", "year","exper",
                        "married","union","wage","black","NorthEast")]
formul = formula(union ~  exper + married + black)

## compute at all periods plus average with the quick method.
output <- felogit( formul, data= sub,Option="quick",compute_T="all")
summary_felogit(output)
```

The table produced by `summary_felogit()` in the R viewer

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

| Variable | Point Est. | se(Point Est.) | pvalue |
|----------|-----------|----------------|--------|
| exper    | -0.0612   | 0.0325         | 0.0594 |
| married  | 0.1595    | 0.2039         | 0.434  |

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

| Period  | Variable | AME/ATE | Estimate | 95% CI |
|---------|----------|---------|----------|--------|
| T=1     | exper    | AME     | -0.0049  | [-0.0102,5e-04] |
|         | married  | ATE     | 0.019    | [-0.0082,0.0461] |
| T=2     | exper    | AME     | -0.005   | [-0.0107,6e-04] |
|         | married  | ATE     | -0.0039  | [-0.0266,0.0187] |
| T=3     | exper    | AME     | -0.0051  | [-0.0108,5e-04] |
|         | married  | ATE     | 0.0238   | [0.0021,0.0456] |
| T=4     | exper    | AME     | -0.0051  | [-0.0106,4e-04] |
|         | married  | ATE     | -0.0098  | [-0.029,0.0094] |
| T=5     | exper    | AME     | -0.005   | [-0.0102,3e-04] |
|         | married  | ATE     | 0.0211   | [6e-04,0.0415] |
| T=6     | exper    | AME     | -0.0048  | [-0.0098,2e-04] |
|         | married  | ATE     | 0.0308   | [-0.0088,0.0704] |
| Average | exper    | AME     | -0.005   | [-0.0104,4e-04] |
|         | married  | ATE     | 0.0135   | [0.0024,0.0245] |

Formula : union ~ exper + married + black.
Nb of discarded individuals: 0.
Nb of observed individuals: 545.
Maximal number of observed periods: 6.

```
The method used to compute 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.
```

When printed in html format it gives reported we obtain Figure 1.

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

| Period | Variable | AME/ATE | Estimate | 95% CI |
|---|---|---|---|---|
| T=1 | exper | AME | -0.0049 | [-0.0102,5e-04] |
| | married | ATE | 0.019 | [-0.0082,0.0461] |
| T=2 | exper | AME | -0.005 | [-0.0107,6e-04] |
| | married | ATE | -0.0039 | [-0.0266,0.0187] |
| T=3 | exper | AME | -0.0051 | [-0.0108,5e-04] |
| | married | ATE | 0.0238 | [0.0021,0.0456] |
| T=4 | exper | AME | -0.0051 | [-0.0106,4e-04] |
| | married | ATE | -0.0098 | [-0.029,0.0094] |
| T=5 | exper | AME | -0.005 | [-0.0102,3e-04] |
| | married | ATE | 0.0211 | [6e-04,0.0415] |
| T=6 | exper | AME | -0.0048 | [-0.0098,2e-04] |
| | married | ATE | 0.0308 | [-0.0088,0.0704] |
| Average | exper | AME | -0.005 | [-0.0104,4e-04] |
| | married | ATE | 0.0135 | [0.0024,0.0245] |

| Variable | Point Est. | se(Point Est.) | pvalue |
|---|---|---|---|
| exper | -0.0612 | 0.0325 | 0.0594 |
| married | 0.1595 | 0.2039 | 0.434 |

*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 + black.
Nb of discarded individuals: 0.
Nb of observed individuals: 545.
Maximal number of observed periods: 6.
The method used to compute 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. Note that the marginal effects in a linear panel data model can also be computed from the packages plm and margins using respectively the functions `plm()` and `margins()`.

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

# References

Davezies L, D'Haultfoeuille X, Laage L (2021). "Identification and Estimation of average marginal 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.

**Affiliation:**

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

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

Christophe Gaillac
CREST
5, avenue Henry Le Chatelier
91 120 Palaiseau, France
E-mail: christophe.gaillac@ensae.fr


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