

The enclosed data and code replicate all tables and graphs in “*Migration, families, and counterfactual families*” by Simone Bertoli, David McKenzie and Elie Murard. Please cite this paper if you use this code.

1) Data sources and variables

The paper uses two data sources, that are both publicly available:

- The American Community Survey (ACS) that we have downloaded from IPUMS USA at <https://usa.ipums.org/usa/acs.shtml>
- The Encuesta Nacional de Ocupación y Empleo (ENOE), which is the Mexican Labor Force Survey that we downloaded from INEGI at <https://www.inegi.org.mx/programas/enoe/15ymas>

We use the 2017-2021 5-year sample of the ACS to construct the final dataset **ACS2017-2021_finalready.dta**, which includes individual observations for foreign-born migrants aged 18 to 60 who migrated from a developing country to the U.S., and moved when they were adults between 18 and 40.

We use the yearly rounds of the ACS between 2008 to 2021 to construct the final datasets **ACS_collapsed_2008_2021.dta** which includes the aggregate numbers of foreign-born migrants (using sample weights) by year of arrival, marital status upon arrival, years spent in the US, and gender.

The original variables we used in the raw data of the ACS are the following:

variable name	storage type	variable label
year	int	Census year
multyear	int	Actual year of survey, multi-year ACS/PRCS
sample	long	IPUMS sample identifier
serial	double	Household serial number
cbserial	double	Original Census Bureau household serial number
numprec	byte	Number of person records following
subsamp	byte	Subsample number
cluster	double	Household cluster for variance estimation
strata	double	Household strata for variance estimation
pernum	int	Person number in sample unit
perwt	double	Person weight
sploc	byte	Spouse's location in household
related	int	Relationship to household head [detailed version]
sex	byte	Sex
age	int	Age
birthqtr	byte	Quarter of birth
marst	byte	Marital status
birthyr	int	Year of birth
marrno	byte	Times married
marrinyr	byte	Married within the past year

yrmarr	int	Year married
bpl	int	Birthplace [general version]
bpld	long	Birthplace [detailed version]
yrimmig	int	Year of immigration
yrsusa1	byte	Years in the United States
educ	byte	Educational attainment [general version]
educd	int	Educational attainment [detailed version]

We use data from the 2005Q1 to the 2018Q4 rounds of the ENOE to construct the final dataset **ENOE_2005_2018.dta**, which includes individual observations for Mexican men aged 18 to 40. The ENOE is a rotating panel that interviews households for five consecutive quarters, which allows observing the occurrence of a migration episode from variations in the household roster across interviews. The dataset **ENOE_2005_2018.dta** only includes the first interview to examine how the marital status at the first interview influences migration decisions that can happen until the 5th interview.

The original variables we used in the raw data of the ENOE are the following:

variable name	storage type	variable label
loc	str4	Localidad
mun	str3	Municipio
t_loc	double	Tamamo de la localidad
cd_a	str2	Ciudad autorrepresentada
ent	byte	Entidad
con	str5	Control, numero consecutivo
d_sem	str3	Distribucion semanal
n_pro_viv	str3	Numero progresivo de vivienda
v_sel	str2	Vivienda seleccionada
n_hog	str1	Numero de hogar
h_mud	str1	Hogar mudado
n_ent	double	Numero de entrevista
per	str3	Periodo de levantamiento
n_ren	str2	Numero de renglon
c_res	double	Condicion de residencia
par_c	str3	Clave de parentesco
sex	double	Sexo
eda	double	Edad
nac_dia	double	Dia de nacimiento
nac_mes	double	Mes de nacimiento
nac_anio	double	Año de nacimiento
l_nac_c	double	Clave de lugar de nacimiento
cs_p12	double	Pregunta 12 ¿Sabe leer y escribir un recado ... ?

cs_p13_1	double	Pregunta 13 ¿Hasta qué grado aprobó ... en la escuela?
cs_p13_2	double	Pregunta 13 ¿Hasta qué año aprobó ... en la escuela?
cs_p17	double	¿ ... Asiste actualmente a la escuela?
e_con	double	Estado conyugal
cs_ad_mot	double	Motivos de la ausencia
cs_ad_des	double	¿ A qué estado de la república o país se fue ... ?
cs_nr_mot	double	¿Cuál es el motivo principal por el que llegó ... ?
cs_nr_ori	double	¿De qué estado de la República Mexicana o país vino ...?
fac	double	Factor de expansión

For the details of the construction of the variables and of the final datasets, please see the manuscript. For further details and questions, please do not hesitate to contact the authors.

2) Description of programs and codes

The replication programs were coded using Stata, version 16. In the main do-file **main_dofile.do**, the user needs to modify the pathnames to match the user's file system: one folder should contain the dataset and another folder should be created to store the output (Tables and Figures). Also, as indicated in **main_dofile.do**, users need to run the ado-file **outreg.ado** in order to load the Stata command “outreg” used to generate the Tables. **main_dofile.do** generates all Tables and Figures included in the paper.

The following codes use the dataset **ACS2017-2021_finalready.dta**:

- **Table1.do** reproduces Table 1 in the main text.
- **Table2.do** reproduces Table 2 in the main text.
- **Table A1-A.5.do** reproduces Tables A.1, A.2, A.3, A.4, A.5 in the appendix.
- **Figure1_figureA.2.do** reproduces Figure 1 in the main text and Figure A.2 in the appendix.
- **Figure2_figureA.1.do** reproduces Figure 2 in the main text and Figure A.1 in the appendix.

The following codes use the dataset **ACS_collapsed_2008_2021.dta**:

- **Figure3_Figure4.do** reproduces Figure 3 and Figure 4 in the main text.

The following codes use the dataset **ENOE_2005_2018.dta**:

- **Table3.do** reproduces Table 3 in the main text

Figure/Table #	Program	Dataset
Table 1	Table1.do	ACS2017-2021_finalready.dta
Table 2	Table2.do	ACS2017-2021_finalready.dta

Table 3	Table3.do	ENOE_2005_2018.dta
Figure 1	Figure1_figureA.2.do	ACS2017-2021_finalready.dta
Figure 2	Figure2_figureA.1.do	ACS2017-2021_finalready.dta
Figure 3	Figure3_Figure4.do	ACS_collapsed_2008_2021.dta
Figure 4	Figure3_Figure4.do	ACS_collapsed_2008_2021.dta
Figure A.1	Figure2_figureA.1.do	ACS2017-2021_finalready.dta
Figure A.2	Figure1_figureA.2.do	ACS2017-2021_finalready.dta
Tables A.1, A.2, A.3, A.4, A.5	Table A1-A.5.do	ACS2017-2021_finalready.dta