Identifying gender specific risk factors for income poverty in urban Mexican households

Electronic Supplemental Material

Note S1: Data cleaning process

After merging the data sources and identifying the available relevant variables, the following analysis for each of the covariates were carried out:

1. Plausibility. This process consists of inspecting the data to discover potential incorrect coding or data errors.
2. Outlier detection. To prevent a few unusual observations from influencing the results, extreme values were identified and excluded from the final data. To do this, boxplots for the continuous variables were used.
3. To ensure having only complete cases in the data set, all the observations with at least one missing value were deleted.

Note S2: Data integration process

The process of integrating the ENIGH with the other eight data sources consists of three steps:

1. I use the ENIGH microdata. In this, observations correspond to individual answers given by the respondents to the ENIGH questionnaire, and each of these individual answers contains a variable to uniquely identify the municipality (CVE_MUN) and the state (CVE_ENT) where the respondent lives. These unique identifiers are assigned by the INEGI.

2. Data at the municipal level from the 2015 Intercensal Population Survey (INEGI, 2015b), the 2015 National Census of Municipal and Delegation Governments (INEGI, 2015a), the National Center for Prevention of Disasters (CENAPRED, 2020), the CONEVAL microdata for poverty estimation (CONEVAL, 2018), the National Population Council (CONAPO, 2016) and the human development index developed by the United Nations Development Program (UNDP, 2019) also contain the municipality unique identifier assigned by INEGI, CVE_MUN. Using this CVE_MUN as a common variable among the datasets, first all the data at the municipal level from these sources were merged, before joining them with the ENIGH microdata. This results in a database with a two-dimensional tree-like hierarchical structure, in which the individual observations of the ENIGH microdata (first dimension) are connected to the estimations at the municipal level (second dimension).

3. Finally, the estimations at the state level from the 2015 National Survey of Quality and Governmental Impact (INEGI, 2015c) and the 2016 National Survey on the Dynamics of Household Relationships (INEGI, 2016b), which contain the state unique identifier assigned by INEGI, CVE_ENT, are merged with the database resulting from step 2. This results in a database with a three-dimensional tree-like hierarchical structure, i.e., the ENIGH individual observations (first dimension) are connected to the information at the municipal level (second dimension), and these, in turn, to the state level estimations (third dimension).

Note S3: Metadata

The following are the metadata describing the attributes needed to reuse and understand the database utilized in this paper. The full dataset can be found in the file called "urban_poverty.RData" and is freely available from Figshare at https://figshare.com/s/b2326294b122fab8cb2d.

Variable	Description
Individual-/household-level covariates	
-Head's age:	Age in years of the household head. **Type:** continuous. **Name in the database:** *edad_jefe* **Source:** (INEGI, 2016a)
-Education level:	Degree of formal education level completed by the household head. **Type:** categorical. *"low"* if the maximum completed level by the head is primary education; *"medium"* if the head has minimum secondary education and a maximum of high school; and, *"high"* if the head has completed at least a university degree. **Name in the database:** *educa_jefe* **Source:** INEGI, 2016a
-Marital status:	Marital status of the household head. **Type:** categorical. *"single"*; *"open-union"*; *"married"*; *"separated"*; *"divorced"*; and, *"widowed"*. **Name in the database:** *edo_conyug* **Source:** INEGI, 2016a
-Indigenous origin:	Indigenous self-identification of the household head. **Type:** categorical. *"yes"* if the head self identifies as indigenous; and, *"no"* otherwise. **Name in the database:** *etnia* **Source:** INEGI, 2016a
-Social networks:	Degree of perception of the household head on the easiness to obtain support from social networks in six hypothetical circumstances: need of money, care due to illness, help to get a job, to be accompanied to a medical appointment, collaboration to improve neighborhood conditions, and child care assistance. **Type:** categorical. *"low"* if obtaining support from social networks in the majority of hypothetical situations is perceived by the head as difficult or impossible; *"high"* if obtaining support from social networks in the majority of hypothetical situations is perceived by the head as easy or very easy; and, *"medium"* otherwise. **Name in the database:** *redsoc_grad* **Source:** CONEVAL, 2018
-Credit card	Holding of a credit card by at least one household member. **Type:** categorical. *"yes"* if at least one member holds a credit card; and, *"no"* otherwise. **Name in the database:** *tarjeta*

Variable	Description
	Source: INEGI, 2016a
-Disability:	Reported status of disability (having a developmental delay; a mental illness; and/or difficulties, or limitations performing one or more basic/everyday activities such as moving their arms, moving their legs, walking, seeing, hearing, speaking, bathing, toileting, eating, dressing, and/or learning basic skills or concepts) of the household head. **Type:** categorical. *"yes"* if at least one member holds a credit card; and, *"no"* otherwise. **Name in the database:** *disc* **Source:** INEGI, 2016a
-Type of household:	Type of household based on the number of members, and the relationship between them. **Type:** categorical. *"one-person"* household consisting of only one member (head). *"nuclear"* household consisting of the head, and his/her partner; the head, his/her partner, and their children; the head, and his/her children; the head, and his/her parents; or the head, and his/her siblings. *"extended"* household consisting of the head, his/her nuclear family (in case of having), and at least another member whose kinship tie with at least one of the rest of household members is beyond the nuclear family kinship ties (i.e. aunts, uncles, nephews, nieces, grandparents, grandchildren, and/or cousins). *"other"* household consisting of the head, his/her nuclear family (in case of having), and/or his/her extended family (in case of having), and at least another member without kinship tie with any of the rest of household members. **Name in the database:** *clase_hog* **Source:** INEGI, 2016a
-Access to food:	Reported status of the access to nutritious and quality food. The respondent is asked if in the last three months, due to lack of money or lack of other resources, at least one of the household members aged 18 or older experienced the following six circumstances: had a diet based on a very small variety of foods; stopped having breakfast, lunch or dinner; ate less than he/she considers should eat; was left without any food; felt hungry but did not eat; and/or ate just once a day or stopped eating for a whole day. Households having at least one member aged under 18 are asked the same questions to separately capture the information for this particular age group. **Type:** categorical. *"yes"* a household having no members aged under 18 is considered having access to nutritious and quality food if the respondent answered affirmatively to less than three of the six questions made (i.e. less than three circumstances experienced in the last three months). A household having at least one member aged under 18 is considered having access to nutritious and quality food if the respondent answered affirmatively to less than four of the 12 questions made; and, *"no"* otherwise. **Name in the database:** *ic_ali* **Source:** CONEVAL, 2018
-Access to health services:	Reported status of the access to public health services. **Type:** categorical. *"yes"* if the head is ascribed or affiliated directly or by kinship to one of the public health institutions or programs; and, *"no"* otherwise. **Name in the database:** *ic_asalud* **Source:** CONEVAL, 2018

Variable	Description
-Dwelling with adequate quality and sufficient space:	Reported status of the access to a dwelling with adequate quality and sufficient space. This indicator takes into account four dwelling's conditions: if the floor is made of concrete or is coated; if the roofs are made of concrete slab or slab joists with roof, wood, metal sheet, asbestos, or any superior quality; if the walls are made of concrete, brick, block, stone, or any superior quality; and/or, if the number of household members per room (including the kitchen, but excluding hallways and bathrooms) is at most 2.5. **Type:** categorical. *"yes"* a household is considered having a dwelling with adequate quality and sufficient space if the dwelling meets the four conditions abovementioned; and, *"no"* otherwise. **Name in the database:** *ic_cv* **Source:** CONEVAL, 2018
-Educational lag:	Reported status of the educational lag of the head. This variable indicates if the head is lagging behind the compulsory level of education according to his/her age. **Type:** categorical. *"yes"* a head has an educational lag if he/she was born before 1982 and has not yet completed the elementary school level; or, if he/she was born on or after 1982 and has not yet completed the secondary level school; and, *"no"* otherwise. **Name in the database:** *ic_rezedu* **Source:** CONEVAL, 2018
-Access to basic housing services:	Reported status of the household access to basic services. This indicator takes into account four basic services: piped water within the dwelling (or outside, but within the dwelling grounds); drainage connected to the public service (or to a septic tank); electricity; and, use of natural or LP gas, or electricity as cooking fuel (or coal but having a chimney). **Type:** categorical. *"yes"* a household is considered having access to basic services if the dwelling has access to the four services abovementioned; and, *"no"* otherwise. **Name in the database:** *ic_sbv* **Source:** CONEVAL, 2018
-Access to social security:	Reported status of the access to social security of the head. This indicator takes into account four circumstances: if the head is economically active and has access to social security (public health services and to the pension system); if the head is not economically active but has access to social security due to direct kinship; if the head is retired and receives a pension; and/or, if the head is 65-years old or older and receives a monetary transfer from a public program. **Type:** categorical. *"yes"* if according to his/her age, working condition, and kinship, the head has access to the corresponding benefits from the social security; and, *"no"* otherwise. **Name in the database:** *ic_segsoc* **Source:** CONEVAL, 2018
-Weekly housework hours:	Time in hours spent on housework by the household head per week. **Type:** continuous. **Name in the database:** *htqueh* **Source:** INEGI, 2016a
Community-level covariates	

Variable	Description
-Social marginalization:	Degree of social marginalization in 2015 of the Municipality of household residence. This indicators takes into account nine socioeconomic indicators at the Municipal level: percentage of the population aged 15 years and over who are illiterate; percentage of the population aged 15 years and over who have not completed elementary school; percentage of the population living in dwellings without drainage nor toilet; percentage of the population living in dwellings without electricity; percentage of the population living in dwellings without piped water; percentage of the population living in overcrowding conditions (number of household members per room, including the kitchen, but excluding hallways and bathrooms, is greater than 2.5); percentage of the population living in dwellings with dirt floor; percentage of the population living in settlements with fewer than 5000 inhabitants; and, percentage of theemployed population having an income of up to two minimum wages. The official methodology elaborated by CONAPO applies the principal component analysis to the data and reduces their dimensionality to a single variable, which is then categorized. **Type:** categorical. *"very low"* *"low"* *"medium"* *"high"* *"very high"* **Name in the database:** *Marg15* **Source:** CONAPO, 2016
-Emergencies due to weather:	Average annual number of declarations of emergency, disaster or contingency due to weather between 2010 and 2015 in the Municipality of household residence. **Type:** continuous. **Name in the database:** *weather* **Source:** CentroNacionaldePrevenciondeDesastres.2020
-Gini index:	Gini index in 2015 of the Municipality of household residence. **Type:** continuous. **Name in the database:** *gini15* **Source:** CONEVAL, 2018
-Human development index:	Human development index in 2015 of the Municipality of household residence. **Type:** continuous. **Name in the database:** *IDH2015* **Source:** UNDP, 2019
-Municipal functional capacities:	Local functional capacities index in 2015 of the Municipality of household residence. This is a composite indicator taking into account five functional capacities of the municipal public administration: capacity to involve relevant stakeholders; capacity to diagnose; capacity to formulate public policies and strategies; capacity to budget, manage, and implement; and, capacity to evaluate. **Type:** continuous. **Name in the database:** *ICFM* **Source:** UNDP, 2019
-Women-to-men ratio of housework hours:	Number of hours spent by women aged 12 years and over doing housework per hour spent by men aged 12 and over doing housework in 2015 in the Municipality of household residence. **Type:** continuous. **Name in the database:** *thnorem* **Source:** INEGI, 2016a
-Women's political participation:	Share of senior positions in the local public administration held by women in 2015 in the Municipality of household residence. **Type:** continuous.

Variable	Description
	Name in the database: *ParPolF* **Source:** INEGI, 2015c
-Migration of women:	Percentage of the 2015 women's population aged 5 years and over in the Municipality of household residence who lived in another State or country in 2010. **Type:** continuous. **Name in the database:** *pres2010_f* **Source:** INEGI, 2015b
-Migration of men:	Percentage of the 2015 men's population aged 5 years and over in the Municipality of household residence who lived in another State or country in 2010. **Type:** continuous. **Name in the database:** *pres2010_m* **Source:** INEGI, 2015b
-Women's household headship:	Percentage of the 2015 population living in women-headed households in the Municipality of household residence. **Type:** continuous. **Name in the database:** *phogjef_f* **Source:** INEGI, 2015b
-Women's economically active population:	Percentage of the 2015 women's population aged 12 years and over who were economically active in the Municipality of household residence. **Type:** continuous. **Name in the database:** *pea_f* **Source:** INEGI, 2015b
-Men's economically active population:	Percentage of the 2015 men's population aged 12 years and over who were economically active in the Municipality of household residence. **Type:** continuous. **Name in the database:** *pea_m* **Source:** INEGI, 2015b
-Women working in the primary sector:	Percentage of the 2015 women's working population aged 12 years and over who were employed in the primary sector in the Municipality of household residence. **Type:** continuous. **Name in the database:** *primario_f* **Source:** INEGI, 2015b
-Men working in the primary sector:	Percentage of the 2015 men's working population aged 12 years and over who were employed in the primary sector in the Municipality of household residence. **Type:** continuous. **Name in the database:** *primario_m* **Source:** INEGI, 2015b
-Women working in the secondary sector:	Percentage of the 2015 women's working population aged 12 years and over who were employed in the secondary sector in the Municipality of household residence. **Type:** continuous. **Name in the database:** *secundario_f* **Source:** INEGI, 2015b
-Men working in the secondary sector:	Percentage of the 2015 men's working population aged 12 years and over who were employed in the secondary sector in the Municipality of household residence. **Type:** continuous. **Name in the database:** *secundario_m* **Source:** INEGI, 2015b
-Women working in the trade sector:	Percentage of the 2015 women's working population aged 12 years and over who were employed in the trade sector in the Municipality of household residence. **Type:** continuous. **Name in the database:** *secundario_f* **Source:** INEGI, 2015b

Variable	Description
-Men working in the trade sector:	Percentage of the 2015 men's working population aged 12 years and over who were employed in the trade sector in the Municipality of household residence. **Type:** continuous. **Name in the database:** *secundario_m* **Source:** INEGI, 2015b
-Women working in the service sector:	Percentage of the 2015 women's working population aged 12 years and over who were employed in the service sector in the Municipality of household residence. **Type:** continuous. **Name in the database:** *servicios_f* **Source:** INEGI, 2015b
-Men working in the service sector:	Percentage of the 2015 men's working population aged 12 years and over who were employed in the service sector in the Municipality of household residence. **Type:** continuous. **Name in the database:** *servicios_m* **Source:** INEGI, 2015b
Regional-level covariates	
-Corruption:	Percentage of the 2015 population aged 18 years and over who considered corruption as a common or very common problem in their State of residence. **Type:** continuous. **Name in the database:** *cor15* **Source:** INEGI, 2015c
-Satisfaction with public services:	Percentage of the 2015 population aged 18 years and over who were satisfied with the basic and on-demand public services provided in their State. **Type:** continuous. **Name in the database:** *satis15* **Source:** INEGI, 2015c
-Violence against women and girls in the community:	Percentage of the 2016 women's population aged 15 years and over who were victims of psychological, physical, and/or sexual gender-based violence at the community level during the last 12 months (between October 2015 and October 2016) in the State of household residence. **Type:** continuous. **Name in the database:** *TPrevCom12Mes* **Source:** INEGI, 2016b
-Violence against women and girls at school:	Percentage of the 2016 women's population aged 15 years and over who were victims of psychological, physical, and/or sexual gender-based violence at school during the last 12 months (between October 2015 and October 2016) in the State of household residence. **Type:** continuous. **Name in the database:** *TPrevEsc12Mes* **Source:** INEGI, 2016b
-Violence against women and girls in the workplace:	Percentage of the 2016 women's population aged 15 years and over who were victims of psychological, physical, and/or sexual gender-based violence in the workplace during the last 12 months (between October 2015 and October 2016) in the State of household residence. **Type:** continuous. **Name in the database:** *TPrevLab12Mes* **Source:** INEGI, 2016b
-Violence against women and girls by an intimate partner:	Percentage of the 2016 women's population aged 15 years and over who were victims of economic, psychological, physical, and/or sexual gender-based violence by an intimate partner during the last 12 months (between October 2015 and October 2016) in the State of household residence. **Type:** continuous.

Variable	Description
	Name in the database: *TPrevRel12Mes* **Source:** INEGI, 2016b
-Violence against women and girls in the family context:	Percentage of the 2016 women's population aged 15 years and over who were victims of economic, psychological, physical, and/or sexual gender-based violence in the family context during the last 12 months (between October 2015 and October 2016) in the State of household residence. **Type:** continuous. **Name in the database:** *TPrevRel12Mes* **Source:** INEGI, 2016b

Note S4: Summary statistics of the response variable

	Man (N=22570)					Woman (N=10503)				
	Min	Max	Mean	Median	SD	Min	Max	Mean	Median	SD
Income-to-poverty ratio	0.00	10.95	1.71	1.23	1.53	0.00	9.41	1.59	1.17	1.36

Note S5: Summary statistics of continuous covariates

	Man (N=22570)					Woman (N=10503)				
	Min	Max	Mean	Median	SD	Min	Max	Mean	Median	SD
Age	21.00	84.00	47.26	46.00	14.36	21.00	88.00	51.76	51.00	15.27
Weekly housework hours	0.00	30.00	4.86	3.00	6.08	0.00	70.00	18.27	14.00	13.62
Emergencies due to weather	0.00	3.17	1.00	0.83	0.83	0.00	3.17	1.02	0.83	0.84
Human development index	0.62	0.94	0.79	0.80	0.05	0.64	0.92	0.79	0.80	0.05
Municipal functional capacities	0.00	0.86	0.39	0.39	0.19	0.00	0.86	0.40	0.39	0.19
Women's political participation	0.00	0.55	0.21	0.20	0.10	0.00	0.55	0.22	0.21	0.10
Migration of women	0.00	0.16	0.04	0.03	0.03	0.00	0.14	0.04	0.03	0.02
Migration of men	0.01	0.16	0.05	0.04	0.03	0.01	0.15	0.05	0.04	0.03
Women's household headship	0.15	0.39	0.27	0.27	0.04	0.16	0.39	0.27	0.28	0.04
Gini index	0.33	0.47	0.39	0.39	0.02	0.33	0.46	0.39	0.39	0.02
Women's economically active population	0.14	0.52	0.37	0.39	0.07	0.12	0.52	0.37	0.39	0.07
Men's economically active population	0.52	0.81	0.70	0.70	0.04	0.53	0.81	0.70	0.70	0.03
Women-to-men ratio of housework hours	1.20	2.73	1.48	1.41	0.24	1.20	2.63	1.47	1.40	0.23
Women working in the primary sector	0.00	0.26	0.02	0.01	0.03	0.00	0.22	0.02	0.01	0.03
Men working in the primary sector	0.00	0.69	0.12	0.06	0.14	0.00	0.64	0.11	0.06	0.14
Women working in the secondary sector	0.03	0.48	0.16	0.14	0.10	0.03	0.48	0.16	0.14	0.10
Men working in the secondary sector	0.11	0.59	0.32	0.31	0.10	0.10	0.62	0.32	0.30	0.10
Women working in the trade sector	0.12	0.32	0.22	0.22	0.04	0.14	0.32	0.23	0.22	0.04
Men working in the trade sector	0.05	0.25	0.16	0.17	0.04	0.05	0.25	0.16	0.17	0.04
Women working in the service sector	0.31	0.81	0.57	0.60	0.09	0.31	0.81	0.58	0.60	0.09
Men working in the service sector	0.11	0.71	0.39	0.41	0.11	0.13	0.66	0.40	0.41	0.11
Corruption	0.73	0.95	0.86	0.86	0.05	0.73	0.95	0.87	0.86	0.05
Satisfaction with public services	0.24	0.54	0.41	0.42	0.08	0.24	0.54	0.41	0.40	0.08
VAWG in the public sphere	0.14	0.37	0.21	0.21	0.05	0.14	0.37	0.21	0.21	0.05
VAWG at school	0.10	0.21	0.17	0.17	0.02	0.10	0.21	0.17	0.17	0.02
VAWG in the family context	0.07	0.13	0.10	0.10	0.02	0.07	0.13	0.10	0.10	0.02
VAWG in the workplace	0.16	0.30	0.22	0.22	0.04	0.16	0.30	0.22	0.22	0.04
VAWG by an intimate partner	0.18	0.33	0.25	0.25	0.04	0.18	0.33	0.25	0.25	0.03

Note S6: Summary statistics of categorical covariates

		Man (N=22570)		Woman (N=10503)	
		N	Pct.	N	Pct.
Education level	low	7454	33.0	4702	44.8
	medium	11434	50.7	4569	43.5
	high	3682	16.3	1232	11.7
Marital status	single	1283	5.7	1593	15.2
	open union	4659	20.6	955	9.1
	married	14513	64.3	1428	13.6
	separated	1030	4.6	2591	24.7
	divorced	435	1.9	1118	10.6
	widowed	650	2.9	2818	26.8
Indigenous origin	yes	6371	28.2	2917	27.8
	no	16199	71.8	7586	72.2
Social networks	low	11285	50.0	6506	61.9
	medium	2868	12.7	1032	9.8
	high	8417	37.3	2965	28.2
Credit card	yes	7455	33.0	2847	27.1
	no	15115	67.0	7656	72.9
Disability	yes	2287	10.1	1719	16.4
	no	20283	89.9	8784	83.6
Type of household	nuclear	15610	69.2	5103	48.6
	one-person	2216	9.8	1893	18.0
	extended	4511	20.0	3360	32.0
	other	233	1.0	147	1.4
Access to food	yes	3806	16.9	2337	22.3
	no	18764	83.1	8166	77.7
Access to health services	yes	3317	14.7	1228	11.7
	no	19253	85.3	9275	88.3
Dwelling with adequate quality and sufficient space	yes	1495	6.6	677	6.4
	no	21075	93.4	9826	93.6
Educational lag	yes	3899	17.3	2585	24.6
	no	18671	82.7	7918	75.4
Access to basic housing services	yes	1866	8.3	811	7.7
	no	20704	91.7	9692	92.3
Access to social security	yes	9720	43.1	4233	40.3
	no	12850	56.9	6270	59.7
Social marginalization	very low	16740	74.2	7848	74.7
	low	3769	16.7	1772	16.9
	medium	1426	6.3	623	5.9
	high	607	2.7	257	2.4
	very high	28	0.1	3	0.0

Note S7: R Code for model estimation

```
######## Code for the replication of estimations in the paper:
######## Identifying gender-specific risk factors for income
######## poverty in urban Mexican households

### Packages ###
if(!require("mboost")) install.packages("mboost")
if(!require("parallel")) install.packages("parallel")

### Database ###
# To be downloaded from https://figshare.com/s/b2326294b122fab8cb2d
load("urban_poverty.RData") # Variables are already zero-centered

### Model ###
fPoverty <- ipov ~

  bols(intercept, intercept = FALSE) +

  # bols(educa_jefelow, intercept = FALSE, df = 1) +
  bols(educa_jefemedium, intercept = FALSE, df = 1) +
  bols(educa_jefehigh, intercept = FALSE, df = 1) +

  # bols(edad_jefe, by = educa_jefelow, intercept = FALSE) +  #
  bols(edad_jefe, by = educa_jefemedium, intercept = FALSE) +  #
  bbs(edad_jefe, by = educa_jefemedium, center = TRUE, df = 1, knots = 20) +
  bols(edad_jefe, by = educa_jefehigh, intercept = FALSE) +  #
  bbs(edad_jefe, by = educa_jefehigh, center = TRUE, df = 1, knots = 20) +

  bols(etnia, intercept = FALSE, df = 1) +

  bols(redsoc_grad, intercept = FALSE, df = 1) +

  # bols(edo_conyugsingle, intercept = FALSE, df = 1) +
  bols(edo_conyugopenunion, intercept = FALSE, df = 1) +
  bols(edo_conyugmarried, intercept = FALSE, df = 1) +
  bols(edo_conyugseparated, intercept = FALSE, df = 1) +
  bols(edo_conyugdivorced, intercept = FALSE, df = 1) +
  bols(edo_conyugwidowed, intercept = FALSE, df = 1) +

  # bols(edad_jefe, by = edo_conyugsingle, intercept = FALSE) +
  # bbs(edad_jefe, by = edo_conyugsingle, center = TRUE, df = 1, knots = 20) +
  bols(edad_jefe, by = edo_conyugopenunion, intercept = FALSE) +
  bbs(edad_jefe, by = edo_conyugopenunion, center = TRUE, df = 1, knots = 20) +
  bols(edad_jefe, by = edo_conyugmarried, intercept = FALSE) +
  bbs(edad_jefe, by = edo_conyugmarried, center = TRUE, df = 1, knots = 20) +
  bols(edad_jefe, by = edo_conyugseparated, intercept = FALSE) +
  bbs(edad_jefe, by = edo_conyugseparated, center = TRUE, df = 1, knots = 20) +
  bols(edad_jefe, by = edo_conyugdivorced, intercept = FALSE) +
  bbs(edad_jefe, by = edo_conyugdivorced, center = TRUE, df = 1, knots = 20) +
  bols(edad_jefe, by = edo_conyugwidowed, intercept = FALSE) +
  bbs(edad_jefe, by = edo_conyugwidowed, center = TRUE, df = 1, knots = 20) +
```

```
  bols(tarjeta, intercept = FALSE, df = 1) +

  bols(disc, intercept = FALSE, df = 1) +

  bols(edad_jefe, intercept = FALSE) +
  bbs(edad_jefe, center = TRUE, df = 1, knots = 20) +

  bols(htqueh, intercept = FALSE) +
  bbs(htqueh, center = TRUE, df = 1, knots = 20) +

  bols(clase_hog, intercept = FALSE, df = 1) +

  bols(ic_ali, intercept = FALSE, df = 1) +
  bols(ic_asalud, intercept = FALSE, df = 1) +
  bols(ic_cv, intercept = FALSE, df = 1) +
  bols(ic_rezedu, intercept = FALSE, df = 1) +
  bols(ic_sbv, intercept = FALSE, df = 1) +
  bols(ic_segsoc, intercept = FALSE, df = 1) +

  brandom(cvegeo, df = 1) +

  bspatial(x, y, center = TRUE, df = 1, differences = 1, knots = 20) +

  bols(weather, intercept = FALSE) +
  bbs(weather, center = TRUE, df = 1, knots = 20) +

  bols(IDH2015, intercept = FALSE) +
  bbs(IDH2015, center = TRUE, df = 1, knots = 20) +

  bols(ICFM, intercept = FALSE) +
  bbs(ICFM, center = TRUE, df = 1, knots = 20) +

  bols(ParPolF, intercept = FALSE) +
  bbs(ParPolF, center = TRUE, df = 1, knots = 20) +

  bols(Marg15, intercept = FALSE, df = 1) +

  bols(pres2010_f, intercept = FALSE) +
  bbs(pres2010_f, center = TRUE, df = 1, knots = 20) +

  bols(pres2010_m, intercept = FALSE) +
  bbs(pres2010_m, center = TRUE, df = 1, knots = 20) +

  bols(phogjef_f, intercept = FALSE) +
  bbs(phogjef_f, center = TRUE, df = 1, knots = 20) +

  bols(gini15, intercept = FALSE) +
  bbs(gini15, center = TRUE, df = 1, knots = 20) +

  bols(pea_f, intercept = FALSE) +
  bbs(pea_f, center = TRUE, df = 1, knots = 20) +

  bols(pea_m, intercept = FALSE) +
```

```
  bbs(pea_m, center = TRUE, df = 1, knots = 20) +

  bols(thnorem, intercept = FALSE) +
  bbs(thnorem, center = TRUE, df = 1, knots = 20) +

  brandom(cveent, df = 1) +

  bols(cor15, intercept = FALSE) +
  bbs(cor15, center = TRUE, df = 1, knots = 20) +

  bols(satis15, intercept = FALSE) +
  bbs(satis15, center = TRUE, df = 1, knots = 20) +

  bols(primario_f, intercept = FALSE) +
  bbs(primario_f, center = TRUE, df = 1, knots = 20) +

  bols(primario_m, intercept = FALSE) +
  bbs(primario_m, center = TRUE, df = 1, knots = 20) +

  bols(secundario_f, intercept = FALSE) +
  bbs(secundario_f, center = TRUE, df = 1, knots = 20) +

  bols(secundario_m, intercept = FALSE) +
  bbs(secundario_m, center = TRUE, df = 1, knots = 20) +

  bols(comercio_f, intercept = FALSE) +
  bbs(comercio_f, center = TRUE, df = 1, knots = 20) +

  bols(comercio_m, intercept = FALSE) +
  bbs(comercio_m, center = TRUE, df = 1, knots = 20) +

  bols(servicios_f, intercept = FALSE) +
  bbs(servicios_f, center = TRUE, df = 1, knots = 20) +

  bols(servicios_m, intercept = FALSE) +
  bbs(servicios_m, center = TRUE, df = 1, knots = 20) +

  bols(TPrevCom12Mes, intercept = FALSE) +
  bbs(TPrevCom12Mes, center = TRUE, df = 1, knots = 20) +

  bols(TPrevEsc12Mes, intercept = FALSE) +
  bbs(TPrevEsc12Mes, center = TRUE, df = 1, knots = 20) +

  bols(TPrevLab12Mes, intercept = FALSE) +
  bbs(TPrevLab12Mes, center = TRUE, df = 1, knots = 20) +

  bols(TPrevRel12Mes, intercept = FALSE) +
  bbs(TPrevRel12Mes, center = TRUE, df = 1, knots = 20) +

  bols(TPrevFam12Mes, intercept = FALSE) +
  bbs(TPrevFam12Mes, center = TRUE, df = 1, knots = 20)

### Three-step strategy ###
```

```r
## Functional gradient descent boosting
modelmext <- gamboost(fPoverty,
                      data = mictpcUr,
                      control = boost_control(mstop = 5000, nu = 0.50,
                                              trace = TRUE,
                                              stopintern = TRUE),
                      weights = mictpcUr$factor,
                      family = QuantReg(tau = ecdf(mictpcUr$ipov)(1310.94/2660.40)),
                      offset = quantile(x = mictpcUr$ipov,
                                        prob = ecdf(mictpcUr$ipov)(1310.94/2660.40)))

# Cross-validation
set.seed(1209)
cvmext <- cvrisk(modelmext,
                 folds = cv(model.weights(modelmext),
                            type = "subsampling"),
                 grid = 1:10000,
                 papply = mclapply,
                 mc.cores = parallel::detectCores())
stopmext <- mstop(cvmext)
modelmext[stopmext]

## Stability selection
p <- length(names(coef(modelmext, which = "")))
stabsel_parameters(p = p, q = 20, cutoff = 0.80)
# Stability selection with unimodality assumption
# Cutoff: 0.8; q: 20; PFER (*):  3.542062
# (*) or expected number of low selection probability variables
# PFER (specified upper bound):  1.614764
# PFER corresponds to signif. level 0.0381 (without multiplicity adjustment)

stabmext <- stabsel(modelmext,
                    cutoff = 0.80,
                    q = 20,
                    sampling.type = "SS",
                    mc.cores = parallel::detectCores())

## Pointwise bootstrap confidence intervals
cimext <- confint(modelmext,
                  B = 1000,
                  level = 0.95,
                  B.mstop = 0,
                  papply = mclapply,
                  cvrisk_options = list(mc.cores = parallel::detectCores()))

## Functional gradient descent boosting
modelfext <- gamboost(fPoverty,
                      data = fictpcUr,
                      control = boost_control(mstop = 5000, nu = 0.50,
                                              trace = TRUE,
                                              stopintern = TRUE),
                      weights = fictpcUr$factor,
                      family = QuantReg(tau = ecdf(fictpcUr$ipov)(1310.94/2660.40)),
```

```
                      offset = quantile(x = fictpcUr$ipov,
                                        prob = ecdf(fictpcUr$ipov)(1310.94/2660.40)))

# Cross-validation
set.seed(1209)
cvfext <- cvrisk(modelfext,
                 folds = cv(model.weights(modelfext),
                            type = "subsampling"),
                 grid = 1:10000,
                 papply = mclapply,
                 mc.cores = parallel::detectCores())
stopfext <- mstop(cvfext)
modelfext[stopfext]

## Stability selection
p <- length(names(coef(modelfext, which = "")))
stabsel_parameters(p = p, q = 20, cutoff = 0.80)
# Stability selection with unimodality assumption
# Cutoff: 0.8; q: 10; PFER (*):  1.61
# (*) or expected number of low selection probability variables
# PFER (specified upper bound):  1.614764
# PFER corresponds to signif. level 0.0316 (without multiplicity adjustment)

stabfext <- stabsel(modelfext,
                    cutoff = 0.80,
                    q = 20,
                    sampling.type = "SS",
                    mc.cores = parallel::detectCores())

## Pointwise bootstrap confidence intervals
cifext <- confint(modelfext,
                  B = 1000,
                  level = 0.95,
                  B.mstop = 0,
                  papply = mclapply,
                  cvrisk_options = list(mc.cores = parallel::detectCores()))

## Functional gradient descent boosting
modelmpov <- gamboost(fPoverty,
                      data = mictpcUr,
                      control = boost_control(mstop = 5000, nu = 0.50,
                                              trace = TRUE,
                                              stopintern = TRUE),
                      weights = mictpcUr$factor,
                      family = QuantReg(tau = ecdf(mictpcUr$ipov)(2660.40/2660.40)),
                      offset = quantile(x = mictpcUr$ipov,
                                        prob = ecdf(mictpcUr$ipov)(2660.40/2660.40)))

# Cross-validation
set.seed(1209)
cvmpov <- cvrisk(modelmpov,
                 folds = cv(model.weights(modelmpov),
                            type = "subsampling"),
```

```
                  grid = 1:10000,
                  papply = mclapply,
                  mc.cores = parallel::detectCores())
stopmpov <- mstop(cvmpov)
modelmpov[stopmpov]

## Stability selection
p <- length(names(coef(modelmpov, which = "")))
stabsel_parameters(p = p, q = 20, cutoff = 0.80)
# Stability selection with unimodality assumption
# Cutoff: 0.8; q: 10; PFER (*):  1.61
# (*) or expected number of low selection probability variables
# PFER (specified upper bound):  1.614764
# PFER corresponds to signif. level 0.0316 (without multiplicity adjustment)

stabmpov <- stabsel(modelmpov,
                    cutoff = 0.80,
                    q = 20,
                    sampling.type = "SS",
                    mc.cores = parallel::detectCores())

## Pointwise bootstrap confidence intervals
cimpov <- confint(modelmpov,
                  B = 1000,
                  level = 0.95,
                  B.mstop = 0,
                  papply = mclapply,
                  cvrisk_options = list(mc.cores = parallel::detectCores()))

modelfpov <- gamboost(fPoverty,
                      data = fictpcUr,
                      control = boost_control(mstop = 5000, nu = 0.50,
                                              trace = TRUE,
                                              stopintern = TRUE),
                      weights = fictpcUr$factor,
                      family = QuantReg(tau = ecdf(fictpcUr$ipov)(2660.40/2660.40)),
                      offset = quantile(x = fictpcUr$ipov,
                                        prob = ecdf(fictpcUr$ipov)(2660.40/2660.40)))

# Cross-validation
set.seed(1209)
cvfpov <- cvrisk(modelfpov,
                 folds = cv(model.weights(modelfpov),
                            type = "subsampling"),
                 grid = 1:10000,
                 papply = mclapply,
                 mc.cores = parallel::detectCores())
stopfpov <- mstop(cvfpov)
modelfpov[stopfpov]

## Stability seltion
p <- length(names(coef(modelfpov, which = "")))
stabsel_parameters(p = p, q = 20, cutoff = 0.80)
```

```r
# Stability selection with unimodality assumption
# Cutoff: 0.8; q: 10; PFER (*):  1.61
# (*) or expected number of low selection probability variables
# PFER (specified upper bound):  1.614764
# PFER corresponds to signif. level 0.0316 (without multiplicity adjustment)

stabfpov <- stabsel(modelfpov,
                    cutoff = 0.80,
                    q = 20,
                    sampling.type = "SS",
                    mc.cores = parallel::detectCores())

## Pointwise bootstrap confidence intervals
cifpov <- confint(modelfpov,
                  B = 1000,
                  level = 0.95,
                  B.mstop = 0,
                  papply = mclapply,
                  cvrisk_options = list(mc.cores = parallel::detectCores()))

save(modelfext, cifext, stopfext, stabfext,
     modelmext, cimext, stopmext, stabmext,
     modelfpov, cifpov, stopfpov, stabfpov,
     modelmpov, cimpov, stopmpov, stabmpov, file = "results_poverty.RData")
```