Essays on structured additive regression models with applications in development economics

Dissertation

zur Erlangung des akademischen Grades eines Doktors
an der Wirtschaftswissenschaftlichen Fakultät
der Georg-August-Universität Göttingen

im Promotionsprogramm
Angewandte Statistik und Empirische Methoden

vorgelegt von
Juan Armando Torres Munguía
geboren am 25. November 1986
in Concepción del Oro, Mexiko

Göttingen, Dezember 2022

Erstgutachterin: Professorin Inmaculada Martínez-Zarzoso, PhD
Zweitgutachter: Professor Konstantin Wacker, PhD
Drittgutachter: Professor Thomas Kneib, PhD
Tag der Disputation: 20. Dezember 2022

Versicherung bei Zulassung zur Promotionsprüfung

Ich versichere,

1. dass ich die eingereichte Dissertation **"Essays on structured additive regression models with applications in development economics"** selbstständig angefertigt habe und nicht die Hilfe Dritter in einer dem Prüfungsrecht und wissenschaftlicher Redlichkeit widersprechenden Weise in Anspruch genommen habe,

2. dass ich das Prüfungsrecht einschließlich der wissenschaftlichen Redlichkeit –hierzu gehört die strikte Beachtung des Zitiergebots, so dass die Übernahme fremden Gedankenguts in der Dissertation deutlich gekennzeichnet ist – beachtet habe,

3. dass beim vorliegenden Promotionsverfahren kein Vermittler gegen Entgelt eingeschaltet worden ist sowie im Zusammenhang mit dem Promotionsverfahren und seiner Vorbereitung

    - kein Entgelt gezahlt oder entgeltgleiche Leistungen erbracht worden sind
    - keine Dienste unentgeltlich in Anspruch genommen wurden, die dem Sinn und Zweck eines Prüfungsverfahrens widersprechen

4. dass ich eine entsprechende Promotion nicht anderweitig beantragt und hierbei die eingereichte Dissertation oder Teile daraus vorgelegt habe.

Mir ist bekannt, dass Unwahrheiten hinsichtlich der vorstehenden Versicherung die Zulassung zur Promotionsprüfung ausschließen und im Falle eines späteren Bekanntwerdens die Promotionsprüfung für ungültig erklärt werden oder der Doktorgrad aberkannt werden kann.

am 20. Dezember 2022
Datum Unterschrift

Declaration of the author's contribution made to the thesis

This thesis is comprised of three essays, the contributions of each author are declared in detail.

- Section 3.1 is an original manuscript that has not been published elsewhere. My contributions to this essay are as follows:
  - Conceptualization.
  - Data curation.
  - Formal analysis.
  - Methodology design.
  - Writing R code to implement the methodology.
  - Visualization of results.
  - Writing – original draft.
  - Writing – review and editing.

  Prof. Martínez-Zarzoso assisted with:
  - Formal analysis.
  - Following up on the research progress.
  - Writing – review and editing.
- Section 3.2 is based on Torres Munguía and Martínez-Zarzoso (2022). My contributions to this essay are as follows:
  - Conceptualization.
  - Data curation.
  - Formal analysis.
  - Methodology design.
  - Writing R code to implement the methodology.
  - Visualization of results.
  - Writing – original draft.

  - Writing – review and editing.

  Prof. Martínez-Zarzoso assisted with:

  - Formal analysis.
  - Following up on the research progress.
  - Writing – review and editing.

- Section 3.3 is based on Torres Munguía and Martínez-Zarzoso (2020, 2021). My contributions to this essay are as follows:

  - Conceptualization.
  - Data curation.
  - Formal analysis.
  - Methodology design.
  - Writing R code to implement the methodology.
  - Visualization of results.
  - Writing – original draft.
  - Writing – review and editing.

  Prof. Martínez-Zarzoso assisted with:

  - Formal analysis.
  - Following up on the research progress.
  - Writing – review and editing.

am 20. Dezember 2022
Datum Unterschrift

Danksagung

An dieser Stelle möchte ich meinen besonderen Dank nachstehenden Personen entgegenbringen, ohne deren Mithilfe und Unterstützung die Anfertigung dieser Doktorarbeit niemals zustande gekommen wäre:

- Mein Dank gilt zunächst Prof. Inmaculada Martínez-Zarzoso, meiner Doktormutter, für die Betreuung dieser Promotionsschrift. Die stimulierenden Diskussionen, die vielen Freiheiten bei der Ideenfindung und das Vertrauen, die sie mir während der gesamten Promotion gewährte, beitrug maßgeblich zum Gelingen dieser Arbeit. Insbesondere danke ich ihr für ihre Verständnis, Geduld und die freundlichen Gespräche auf menschlicher und persönlicher Ebene, die werden mir immer als bereichernder Austausch in Erinnerung bleiben.
- Mein Dank gilt des weiteren Prof. Konstantin Wacker für die Anfertigung des Zweitgutachtens, die immer sehr freundliche Zusammenarbeit und die zahlreichen Ideen.
- Weiterhin danke ich Prof. Thomas Kneib für die wissenschaftliche und hilfsbereite Betreuung als Drittgutachter.
- Esta tesis de doctorado no hubiera sido posible sin el apoyo y amor incondicionales de mi familia. Especial dedicatoria a mi padre que desde el cielo me cuida. Gracias por enseñarme el valor de la honestidad y por mostrarme que la vida está hecha para disfrutarse al máximo, en todo momento. Gracias madre por enseñarme el concepto del amor, el valor de la disciplina y el esfuerzo. Gracias a mi hermana y a mi hermano. Rosy y Aldo, ustedes son mi máximo ejemplo en la vida. Gracias familias Narváez Torres y Torres Fernández por siempre acompañarme en mi camino y porque me han dado cinco alegrías especiales: Jimena, Mariana, Ivanna, Jorge y Santiago. Gracias Jorge y gracias Sandra. Gracias también a las familias Torres García, Munguía Gómez, Aguilar Gómez, Narváez Méndez y Fernández Casas.

- Quiero agradecer también a los amigos que he conocido en México y que aún conservo. Gracias por mostrarme que la distancia y el tiempo solo hacen la amistad y cariño mucho más fuertes. Gracias Alexei, Anais, Andrés, Björn, Freddy, José, Julieta, Julio, Luis, Miriam, Nadia y Zack. I also wish to show my gratitude to the great human beings I have met during my PhD studies: Alan, Alyona, Ana, Andreas, Anna Stampa, Anna Wegrzynowicz, Banoo, Brenda, Daniela, Daniel, Dinah, Dominga, Fabio, Felipe, Felix, Giulia Butera, Giulia Trovato, Hooman, Isabel, Isabella, Javiera, Jonathan, Leandra, Lilian, Lucie, Luis, Luisa, Lukas, Maca, Malin, Marvin, Matilde, Max, Michael, Miriam, Momo, Nati, Paul, Rebecca, Ryan, and Ugo. Thanks for being part of this special episode of my life, for the pleasant distraction, the multiple parties, the serenatas, the dances, the songs sang in the streets, the piñatas, the trips, the food, the laughs, the drinks, the talks, the messages, the calls, for taking care of me, for visiting me in my house, but more specially, thanks for showing me what true friendship and love is. Thank you all for being the reason I believe in the goodness of people.

- Mein außerordentlicher Dank gilt Herr Robin Schaeffer, ohne deren professionelle Hilfe in der beschwerlichen Zeit meiner Promotion wäre ich heute nicht die Person, die ich bin. Ich werde niemals vergessen, was ich über Selbstliebe, Selbstwertgefühl und Selbstbewusstsein gelernt habe.

- Tief dankbar bin ich Frau Marty Auer für ihre unglaublich hilfreiche Unterstützung und ihr Verständnis, vor allem aber ihr professioneller Beistand und der menschliche Halt, haben mir Kraft und Mut zur Anfertigung und Vollendung meiner Dissertation gegeben. Danke dass Sie mich daran erinnert haben, wie wichtig die Prinzipien sind und den Stellenwert von Liebe, Respekt und Ruhe.

Contents

List of publications

- Torres Munguía, J. A., & Martínez-Zarzoso, I. (2022). Determinants of emotional intimate partner violence against women and girls with children in mexican households: An ecological framework [PMID: 35135364]. *Journal of Interpersonal Violence*, *0*(0), 08862605211072179. https://doi.org/10.1177/08862605211072179
- Torres Munguía, J. A., & Martínez-Zarzoso, I. (2021). Examining gender inequalities in factors associated with income poverty in mexican rural households. *PloS one*, *16*(11), e0259187. https://doi.org/10.1371/journal.pone.0259187
- Torres Munguía, J. A., & Martínez-Zarzoso, I. (2020). What determines poverty in mexico? a quantile regression approach

List of Figures

List of Tables

Abstract

Structured additive regression models are a particular class of models that provide a flexible framework to deal with a wide class of effects, including linear, nonlinear, random, spatial, and interaction effects, which enables the specification of more complex but more realistic models.

The goal of this dissertation is to use these models to address practical issues in three relevant topics in the field of development economics. First, a Gaussian model is used to study gendered inequalities in time allocation to unpaid housework among partnered women and men. In the second study, we are confronted with the problem of identifying the risk factors associated with emotional intimate partner violence, for which a probit model is used. In the third study, quantile models are applied to examine heterogeneous gendered effects of a set of risk factors associated with the income-to-poverty ratio of the poor and extremely poor families.

Given the complex structure of the models used in the three abovementioned cases, an estimation cannot be computed by traditional inference techniques. To overcome this issue, it is implemented a three-step strategy consisting on the use of the boosting algorithm, complementary pairs stability selection with per-family error rate control, and the calculation of pointwise bootstrap confidence intervals.

From a statistical standpoint, the methodology helps to overcome common issues in regression in development economics, such as dealing with different types of response variables, the inclusion of potential nonlinear (or even *a priori* unknown) effects of continuous covariates on the response, select the relevant variables at their most suitable functional form, dealing with hierarchical data, to account for spatially correlated observations, to introduce complex interaction effects, and to avoid multicollinearity.

From an empirical perspective, the method applied allows to illustrate how the utilization of the structured additive models contributes to enhancing knowledge on these phenomena by providing new relevant insights on the matter. Findings in the three studies not only yield evidence about significant covariates that were either hitherto unknown, understudied, or that have not yet been tested empirically, but they are also relevant for the design of public policies, such as the identification of the relevance of the individual, household, communities, and regional factors in these studies, the existence of age-varying effects, the determination of the circumstances in which women and men face particular disadvantages, and the identification of some specific risk subgroups of the population that are generally overlooked.

Zusammenfassung

Strukturiert additive Regressionsmodelle sind eine bestimmte Klasse von Modellen, die einen flexible Struktur für den Umgang mit verschiedenartigen Kovariableneffekten bietet, einschließlich linearer, nichtlinearer, zufälliger, räumlicher und Interaktionseffekte, was die Spezifikation von komplexerer, aber wirklichkeitsgetreuer Modelle ermöglicht.

Das Ziel dieser Dissertation ist es, diese Modelle zu nutzen, um praktische Fragestellungen in drei relevanten Themenfeldern der Entwicklungsökonomie zu untersuchen. Zunächst wird ein Normalverteilungsmodell verwendet, um geschlechtsspezifische Ungleichheiten bei der Zeitverwendung von Frauen und Männer in Partnerschaft für unbezahlte Hausarbeit. In der zweiten Studie werden wir mit dem Problem konfrontiert, die Risikofaktoren emotionaler Partnergewalt zu identifizieren, für die ein Probit-Modell verwendet wird. In der dritten Studie werden Quantilmodelle angewendet, um heterogene geschlechtsspezifische Auswirkungen einer Reihe von Risikofaktoren zu untersuchen, die mit dem Verhältnis von Einkommen zu Armut in armen und extrem armen Familien verbunden sind.

Angesichts der komplexen Struktur der Modelle, die in den drei oben genannten Fällen verwendet werden, kann eine Schätzung nicht durch herkömmliche Inferenztechniken berechnet werden. Um dieses Problem zu lösen, wird eine dreistufige Strategie implementiert, die aus der Verwendung des Boosting-Algorithmus, der *complementary pairs stability selection* mit *per-family error rate control* und der Berechnung von punktweisen Bootstrap-Konfidenzintervallen besteht.

Aus statistischer Sicht hilft die Methodik dabei, häufige Probleme bei der Regression in der Entwicklungsökonomie zu überwinden, z.B. verschiedene Arten von Zielgrößen, Auswahl der relevanten Variablen in ihrer am besten geeigneten funktionalen Form, Umgang mit hierarchischen Daten und räumlich korrelierte Beobachtungen zur Berücksichtigung, komplexe Interaktionseffekte einzuführen und Multikollinearität zu vermeiden.

Aus empirischer Sicht ermöglicht die angewandte Methode darzustellen, wie die Nutzung der strukturierten additiven Modelle dazu beiträgt, das Wissen über diese Phänomene zu erweitern, indem sie neue relevante Erkenntnisse zu diesem Thema liefern. Die Ergebnisse der drei Studien geben nicht nur Hinweise auf signifikante Kovarianzen, die entweder bisher unbekannt, zu wenig untersucht oder noch nicht empirisch getestet wurden und gleichzeitig für die Gestaltung öffentlicher Maßnahmen relevant sind, beispielsweise für die Ermittlung der Relevanz von Einzel-, Haushalts-, Gemeinde- und regionale Faktoren in diesen Studien, das Vorhandensein von altersabhängigen Effekten, die Bestimmung der Umstände, unter denen Frauen und Männer besonderen Benachteiligungen ausgesetzt sind, und die Identifizierung einiger spezifischer Risikountergruppen der Bevölkerung, die im Allgemeinen übersehen werden.

1. Introduction

"Statistics is the grammar of Science."
Karl Pearson

Regression analysis is one of the most popular statistical tools utilized today by researchers in several fields of science, including development economics. Here, the goal is to identify and describe how a set of covariates $x_1, x_2, ..., x_k$, also known as independent variables, is associated with a variable y of primary interest, called response or dependent variable. Traditionally, this linkage is defined by:

$$y = f(x_1, x_2, ..., x_k) + \varepsilon \tag{1.1}$$

where $f(x_1, x_2, ..., x_k)$ is an unknown function modelling the relationship between y and $x_1, x_2, ..., x_k$, and ε is the error term. In the context of classical linear models, $f(x_1, x_2, ..., x_k)$ is assumed to be a linear combination of the k covariates. Hence, considering $i = 1, ..., n$ data points:

$$y_i = \beta_0 + \beta_1 x_{i1} + ... + \beta_k x_{ik} + \varepsilon_i \tag{1.2}$$

where $\beta_0, ..., \beta_k$ are the unknown regression parameters to be estimated and that indicate the direction and strenght of the covariate effect on the response, and $\varepsilon_1, ..., \varepsilon_n$ are the error terms, which follow a normal distribu-

tion and are independent identically distributed (i.i.d) with $E(\varepsilon_i) = 0$ and $Var(\varepsilon_i) = \sigma^2$. These assumptions about ε_i carry over the dependent variable and therefore $y_i \sim \mathcal{N}(\mu_i, \sigma^2)$, where $\mu_i = \beta_0 + \beta_1 x_{i1} + ... + \beta_k x_{ik}$.

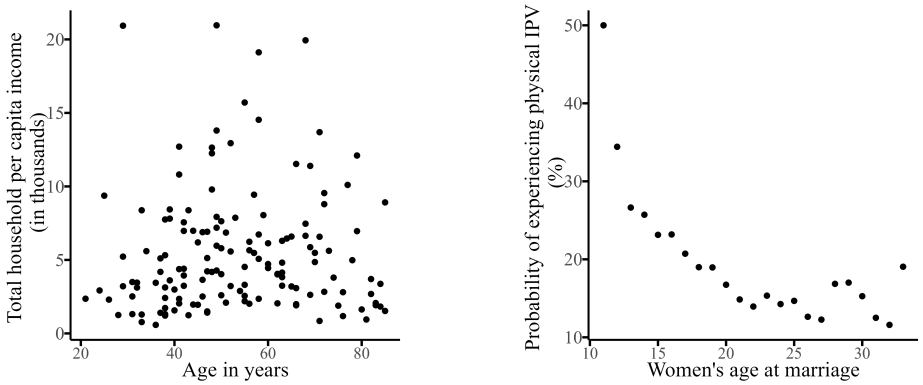
When y_i follows a distribution other than Gaussian but belongs to an exponential family, we define a generalized linear model:

$$h(\mu_i) = \beta_0 + \beta_1 x_{i1} + ... + \beta_k x_{ik} + \varepsilon_i \tag{1.3}$$

where $h(\mu_i)$ is the identity link, a function connecting y_i with the linear component $\beta_0 + \beta_1 x_{i1} + ... + \beta_k x_{ik}$.

Despite the key significance of these approaches, however, in many real world problems in development economics, either by a theoretical basis, a lack of certainty, or absence of prior knowledge, a purely linear effect might not always be suitable nor sufficient for describing the association of the response with the covariates. In order to illustrate this idea, let consider the two motivating examples depicted in Fig. 1.1.

Fig. 1.1 Two examples of nonlinearities in development economics



(a) **(b)**

See implementation details in Supplementary information 5.1

Source: Own elaboration based on data from INEGI (2016a, 2016c).

First, Fig. 1.1a shows a scatter plot of the total household per capita income (in thousands) and age in years of the household head. For simplicity and in order to provide a clear visualization, the information exclusively corresponds to urban families headed by a woman in a given municipality in Mexico (Saltillo, in this case). Broadly speaking, Fig. 1.1a suggests that age has a nonlinear effect on the income, particularly, that this association

is approximately described by an inverted U-shaped curve. The second case is presented in Fig. 1.1b and it shows the relationship in Mexico between women's age at marriage and their probability of ever experiencing physical violence perpetrated by the intimate partner (IPV). As can be seen from Fig. 1.1b, the likelihood of being a victim decreases as the woman's age at marriage increases. The decreasing pattern is particularly clear for women who married as children, and indeed, for those marrying after about age 20 the probability of victimization appears to be stable at around 15 percent.

Within the linear models approach, one of the alternatives to fit nonlinearities, such as the abovementioned cases, is to apply a variable transformation or to introduce polynomials (Harrell Jr., 2015). However, although these alternatives are well documented, easily applied, and relatively straightforward to interpret, it is not always possible to find a transformation or polynomial to render the data suitable for subsequent linear regression given the limited number of potential variable modifications (Fahrmeir et al., 2013).

Another methodological alternative to deal with nonlinear effects is to move away from the traditional linear approach to nonparametric regression models. In these models, the linear component is replaced by a much more flexible part captured by an unspecified smooth function called regression splines (Eilers & Marx, 1996). For simplicity, let covariate x_1 follow a nonlinear relationship with the response variable y:

$$y_i = s(x_{i1}) + \varepsilon_i \tag{1.4}$$

where, similar as in the linear regression case, ε_i are the error terms with $E(\varepsilon_i) = 0$ and $Var(\varepsilon_i) = \sigma^2$. The $s(x_{i1})$ can be represented as a linear model by choosing m known basis functions b_j, with unknown parameters γ_j:

$$s(x_{i1}) = \sum_{j=1}^{m} \gamma_j b_j(x_{i1}) \tag{1.5}$$

Then

$$y_i = \sum_{j=1}^{m} \gamma_j b_j(x_{i1}) + \varepsilon_i \tag{1.6}$$

Nevertheless, there is a price to pay for this flexibility: estimation is practically intractable or computationally infeasible in the framework of high

dimensional data settings (Christensen, 2019). Broadly speaking, high dimensionality occurs when a (very) large number of parameters k relative to the number of observations n, is introduced in a regression model (Johnstone & Titterington, 2009). To formally express this, let extent Equation 1.6 to the high dimensional case including all the possible interactions of covariates:

$$y_i = s(x_{i1}, ..., x_{ik}) + \varepsilon_i = \sum_{j_1=1}^{m_1} ... \sum_{j_k=1}^{m_k} \gamma_{j_1 ... j_k} b_{j_1 ... j_k}(x_{i1}, ..., x_{ik}) + \varepsilon_i \qquad (1.7)$$

In many applications, working with high dimensional data has become growingly recurrent and important. Such data contexts arise as a result of multiple interlinked situations. First, there is a continuously increasing availablity of information to characterize the units of observation in a study. Specifically about development economics, data may come from censuses, administrative records, or household surveys that collect information at various levels: individual, household, community, regional, national, and international. Moreover, as a result of the expanding utilization of high-tech tools researchers on the field have within reach other potential sources producing a plethora of data, such as Geographic Information Systems (GIS) or web-based data (Belloni et al., 2014). This wide availability of data also allows researchers to examine complex and multifaceted phenomena (such as crime, poverty, and inequality) from many different angles by adding multiple covariates on different subjects, including topics such as corruption, climate, social networks, or gender issues, to better characterize and understand the problem of interest.

Given the ubiquity of these high dimensional data settings in research, to overcome their inherent difficulties, Equation 1.7 can be specified as the sum of unknown functions for each of the individual covariate effects:

$$y_i = s_1(x_{i1}) + ... + s_k(x_{ik}) + \varepsilon_i = \sum_{j_1=1}^{m_1} \gamma_{j_1} b_{j_1}(x_{i1}) + ... + \sum_{j_k=1}^{m_k} \gamma_{j_k} b_{j_k}(x_{ik}) + \varepsilon_i \quad (1.8)$$

This Equation corresponds to the general representation of the so-called additive regression models (Hastie & Tibshirani, 1986, 1999). The main advantage of replacing the linear effects of the regression model by this additive structure is that functions $s_1(x_{i1}), ..., s_k(x_{ik})$ can be of different type to deal

with various classes of variables and effects. By introducing nonlinear, linear, random, spatial, and interaction effects into Equation 1.8, the resulting formula is known as a structured additive regression model (Fahrmeir et al., 2013).

In this thesis, the objective is to use these structured additive regression models to address practical issues in three relevant topics in the area of development economics. First, in Section 3.1 we use a structured additive Gaussian model to study gendered inequalities in time allocation to unpaid housework among partnered women and men in Mexico. In this example, we utilize a data set composed of 16,167 observations and 30 potential covariates. In Section 3.2 we are confronted with the problem of identifying the risk factors associated with emotional IPV in Mexico. To that end, we generate a data set with more than 35,000 observations and 39 variables, to which we apply a structured additive probit model. The data set integrates 10 information sources, allowing us to properly characterize the context of IPV from a multilevel perspective, including information about the individuals, their relationship, the community, and the society where they live. Finally, in Section 3.3, we apply structured additive quantile models to a cross-sectional data set containing information on 4,434 women-headed and 14,877 men-headed Mexican households to examine heterogeneous gendered effects of a set of theoretical risk factors on two of the lowest quantiles of the income-to-poverty ratio distribution, namely the corresponding to poor and extremely poor families. For each model, we introduce 42 variables at the individual/household, community, and regional levels.

Given the complex structure of the models used in the three cases and their high dimensionality, an estimation cannot be computed by traditional methods. To overcome this issue, we implement the following three-step strategy (see Section 2.2):

- Step 1: Estimation via component-wise gradient boosting algorithm (see Section 2.2.1).
- Step 2: Stability selection to avoid the erroneous selection of non-relevant variables (see Section 2.2.2)
- Step 3: Finally, 95% pointwise bootstrap confidence intervals are calculated for the subset of effects selected as stable in step 2 (see Section 2.2.3).

From a statistical standpoint, this strategy helps us to overcome eight common issues in regression models in development economics:

- To deal with different types of response variables (continuous, categorical, etc.).
- The inclusion of potential nonlinear (or even *a priori* unknown) effects of continuous covariates on the response.
- To deal with a hierarchical data structure, in which individual observations are connected to the information for the communities, and these, in turn to the regional information.
- To account for spatially correlated observations.
- To introduce interaction effects between a categorical and a continuous covariate.
- To consider interaction effects between two continuous covariates.
- To perform estimation with automatic identification of significant covariates (variable selection) and determination of the functional form of their linkage with the dependent variable (model choice).
- To avoid multicollinearity problems.

From an empirical perspective, the method applied allows to illustrate how the utilization of the structured additive models could contribute to enhancing knowledge on these phenomena by providing new relevant insights on the matter.

The content of this thesis is grouped in five parts. Following this introduction, in Chapter 2 the basis of the structured additive regression models are presented. Then, Chapter 3 presents the three application cases for these models. In Chapter 4 final remarks are commented with a focus on the use of structured additive regression models in development economics, the contribution of this research project on the application studies, and future research. Finally, supplementary information is provided in Chapter 5 to help the reader to better understand, analyze, and replicate all the analysis in this thesis.

2. Structured additive regression models

"Statisticians, like artists, have the bad habit of falling in love with their models."
George Box

Structured additive regression models are a particular class of additive models (Equation 1.8) combining different types of effects, namely linear, nonlinear, spatial, random, and/or interaction effects in a single representation. Let consider the response variable y and a set of p categorical $w_1, ..., w_p$, and q continuous variables $z_1, ..., z_q$. Hence, for $i = 1, ..., n$:

$$y_i = \beta_0 + \beta_1 w_{i1} + ... + \beta_p w_{ip} + s_1(z_{i1}) + ... + s_q(z_{iq}) + \varepsilon_i \qquad (2.1)$$

where β_0 is the constant term for the model intercept, $\beta_1, ..., \beta_p$ are the unknown regression parameters for the effect of the categorical covariates, $s_1(z_{i1}), ..., s_q(z_{iq})$ are smooth functions for the nonlinear effects of the continuous covariates, and $\varepsilon_1, ..., \varepsilon_n$ are the error terms. To avoid the problem of identification inherent to additive models (Hothorn et al., 2020), all $s_1(z_{i1}), ..., s_q(z_{iq})$ are defined as:

$$\sum_{i=1}^{n} s_1(z_{i1}) = ... = \sum_{i=1}^{n} s_q(z_{iq}) = 0 \tag{2.2}$$

Now, incorporating other types and more complex effects into Equation 2.1, we have:

$$\begin{aligned} y_i = \beta_0 + \sum_{l=1}^{p} \beta_l w_{il} + \sum_{r=1}^{q} s_r(z_{ir}) + s_{geo}(lon_i, lat_i) + \\ s_{int_1}(z_{id}) w_{ig} + s_{int_2}(z_{ie}, z_{if}) + \varepsilon_i \end{aligned} \tag{2.3}$$

where $s_{geo}(lon, lat)$ is a component to model spatial effects of geographic coordinates lon and lat, $s_{int_1}(z_d)w_g$ is a component capturing the interaction effect of the continuous covariate z_d and the categorical variable w_g, and $s_{int_2}(z_e, z_f)$ denotes the interaction effect between the continuous covariates z_e and z_d.

Evidently, Equation 2.3 can also be extended to the case of non-normally distributed response variables (count, categorical, or ordered) similarly to the case of generalized linear models in Equation 1.3. Thus, recalling that $E(y_i) = \mu_i$:

$$\begin{aligned} h(\mu_i) = \beta_0 + \sum_{l=1}^{p} \beta_l w_{il} + \sum_{r=1}^{q} s_r(z_{ir}) + s_{geo}(lon_i, lat_i) + \\ s_{int_1}(z_{id}) w_{ig} + s_{int_2}(z_{ie}, z_{if}) + \varepsilon_i \end{aligned} \tag{2.4}$$

Equations 2.3 and 2.4 include all the covariate effects introduced in the models that will be used in the applications developed in this thesis. However, more effects can be added in the context of structured additive regression models (Wood, 2017). In the following section we describe more in detail all the right-hand-side model components from Equations 2.3 and 2.4.

2.1 Model components

As previously mentioned, in the context of structured additive regression models it is dealt with different types of effects according to the various types of independent variables introduced in a model. In the following lines the different types of effects used in this thesis are described.

2.1.1 Parametric effects for categorical variables

In Equations 2.3 and 2.4, the effect for the p categorical variables is captured by $\beta_0 + \sum_{l=1}^{p} \beta_l w_{il}$. This is also known as the parametric part of the model. Let us suppose the variable w has $h \geq 2$ categories. Then, to estimate the effect of w on the response y, $h-1$ dummy variables are specified:

$$w_h = \begin{cases} 1 & \text{if } w = h \\ 0 & \text{otherwise} \end{cases} \tag{2.5}$$

The remaining category works as the reference. To estimate its effect, entries of zeros are introduced in all the $h-1$ dummy variables and therefore the effect of the reference is captured by β_0. Interpretation of the parameters is basically the same as in other regression approaches. Parameters of the $h-1$ dummy variables indicate the difference in the effect of the corresponding category on the response with respect to the effect of the reference category, captured by β_0.

2.1.2 Nonparametric effects for continuous variables

$\sum_{r=1}^{q} s_r(z_{ir})$ is the model component for the q continuous variables, where parameters $s_r(z_{ir})$ are smooth functions based on basic splines or B-splines (Eilers & Marx, 1996). These are expressed as in Equation 1.5 by choosing m known functions b_j, with unknown parameters γ_j. Thus $s_r(z_{ir}) = \sum_{j=1}^{m} \gamma_j b_j(z_{ir})$.

The underlying idea of B-splines is that $s_r(z_{ir})$ can be determined by piecewise polynomials of degree m called splines, that consists of $m+1$ intervals joined at m evenly spaced knots within the domain of z_r, in a $(m-1)$-times continuously differentiable form. Certainly, the splines depend significantly on the number and position of the knots: a very large m may lead to a low bias but a large variance, while a small m can produce a

function having a small variance but a large bias (Wood, 2017). To control this trade-off between smoothness and fit, Eilers and Marx (1996) proposed using a relative large m to achieve enough flexibility and applying difference penalties on the basis coefficients of adjacent B-splines to prevent overfitting and thus achieve smoothness, technique that they called penalized splines or simply P-splines. The representation of this penalty is:

$$\lambda P(\gamma) = \lambda \sum_{j=d+1}^{m} \Delta_d(\gamma_j) = \lambda \gamma' \mathbf{K} \gamma \tag{2.6}$$

where $\lambda \geq 0$ is a smoothing parameter, m is the number of basis functions, Δ_d is a dth-order difference operator, $\mathbf{K}$ is a symmetric, positive semidefinite penalty matrix given by $\mathbf{K} = \mathbf{D}'_{\mathbf{d}}\mathbf{D}_{\mathbf{d}}$, with $\mathbf{D_d}$ being a dth-order difference matrix. It is important to highlight that a too large λ yields to a more flexible effect (bias towards a nonlinear effect). In the framework of a model with multiple continuous covariates it is also important to make their effects comparable in terms of flexibility to avoid biased estimates (Hofner et al., 2016). This can be addressed by having a λ such that $df(\lambda) = 1$. Nevertheless, since a $(d-1)$th-order difference polynomial remains unpenalized, it is not possible to make $df(\lambda)$ arbitrarily small. Therefore, in $\sum_{r=1}^{q} s_r(z_i)$, each of these functions is decomposed into a linear part (unpenalized polynomial) and a nonlinear polynomial (penalized, smooth deviation from the unpenalized polynomial) estimated by P-splines (Hofner et al., 2014; Kneib et al., 2009). This decomposition is key in this context, since it enables us to leave *a priori* the functional form of the relationship between the response and the continuous covariates unspecified. As a consequence, the effect of every $s_r(z_{ir})$ can result in:

- Non-significant covariate effect;
- "purely" linear effect;
- nonlinear effect; or,
- a combined effect of a linear and a nonlinear effects.

In the presence of "purely" linear effects, the estimated parameter indicates the change in the response variable per unit change in the continuous covariate. For nonlinear effects interpretation is best done by visualizing the corresponding estimations.

2.1.3 Spatial effects

Spatial effects are introduced in component $s_{geo}(lon_i, lat_i)$ and are estimated by bivariate tensor product P-splines. A bivariate tensor product basis is applied to represent a smooth function of two continuous covariates, in this case the geographic coordinates, and it is derived by considering all pairwise products of them, yielding to the representation:

$$s_{geo}(lon_i, lat_i) = \sum_{j=1}^{m_{lon}} \sum_{k=1}^{m_{lat}} \gamma_{jk} b_k^{lon}(lon_i) b_j^{lat}(lat_i) \tag{2.7}$$

where m_{lon} and m_{lat} are the number of known basis functions b_j^{lon} and b_j^{lat}, respectively, with unknown parameters γ_{jk}. The penalty for a bivariate tensor basis is constructed in an analogous way to the P-splines in the univariate case (see Equation 2.6):

$$\lambda P(\gamma) = \lambda \gamma' \mathbf{K} \gamma = \lambda \gamma' [(\mathbf{I_{m_{lon}}} \otimes \mathbf{K_{lat}} + \mathbf{I_{m_{lat}}} \otimes \mathbf{K_{lon}})] \gamma \tag{2.8}$$

where $\mathbf{I_{m_{lon}}}$ and $\mathbf{I_{m_{lat}}}$ are the m_{lon}- and m_{lat}-dimensional identity matrices, $\mathbf{K_{lat}}$ and $\mathbf{K_{lon}}$ are symmetric, positive semidefinite penalty matrices, and operator $\otimes$ indicates the Kronocker product.

2.1.4 Interaction effects

Interaction effects occur when the effect of a covariate on the response differs according to the value(s) of one or more other covariates. Here we only explore interactions between two covariates. Hence, two types of interacting effects are considered: the interaction of a continuous independent variable with a categorical covariate (varying effects), and the interaction between two continuous covariates (interaction surfaces).

Varying effects

Component $s_{int_1}(z_{id}) w_{ig}$ in Equations 2.3 and 2.4 denotes the interaction between a continuous and a categorical covariate. Broadly speaking, these parameters capture how the effect of the categorical variable w_g on the response varies over the co-domain of the continuous covariate z_d.

Let us consider the simplest case in which w_g is a binary variable, then $s_f(z_d)$ captures the nonlinear effect of the continuous variable z_d if $w_g = 0$ and $s_d(z_d) + s_{int_1}(z_d) w_g + \beta_f w_d$ represents the effect of z_d when $w_g = 1$.

Interaction surfaces

The $s_{int_2}(z_e, z_f)$ part in Equations 2.3 and 2.4 indicates the interaction between the continuous covariates z_e and z_f. This component is called interaction surface and is estimated by bivariate tensor product P-splines, exactly as in the case of the spatial effects. Consequently,

$$s_{int_2}(z_{ie}, z_{if}) = \sum_{j=1}^{m_{z_e}} \sum_{k=1}^{m_{zf}} \gamma_{jk} b_k^{z_e}(z_{ie}) b_j^{z_f}(z_{if}) \tag{2.9}$$

where m_{z_e} and m_{z_f} are the number of known basis functions $b_j^{z_e}$ and $b_j^{z_f}$, respectively, with unknown parameters γ_{jk}. Therefore, the penalty is:

$$\lambda P(\gamma) = \lambda \gamma' \mathbf{K} \gamma = \lambda \gamma' [(\mathbf{I}_{\mathbf{m_{z_e}}} \otimes \mathbf{K}_{\mathbf{z_f}} + \mathbf{I}_{\mathbf{m_{z_f}}} \otimes \mathbf{K}_{\mathbf{z_e}})] \gamma \tag{2.10}$$

where $\mathbf{I}_{\mathbf{m_{z_e}}}$ and $\mathbf{I}_{\mathbf{m_{z_f}}}$ are the m_{z_e}- and m_{z_f}-dimensional identity matrices, $\mathbf{K}_{\mathbf{z_f}}$ and $\mathbf{K}_{\mathbf{z_e}}$ are symmetric, positive semidefinite penalty matrices, and operator $\otimes$ indicates the Kronocker product.

2.1.5 Random effects

In this thesis, in Sections 3.2 and 3.3, we also consider a hierarchical data structure in which individual observations are connected to the information for the communities, and these, in turn to the regional information. The random effects are introduced to take into account this multilevel structure. Let consider the observations $i = 1, .., n_v$ in clusters $v = 1, ..., m$, then Equation 2.3 can be represented as:

$$\begin{aligned} y_{vi} = \beta_{0v} + \sum_{l=1}^{p} \beta_l w_{vil} + \sum_{r=1}^{q} s_r(z_{vir}) + s_{geo}(lon_{vi}, lat_{vi}) + \\ s_{int_1}(z_{vid}) w_{vig} + s_{int_2}(z_{vie}, z_{vif}) + \phi_{0v} + \varepsilon_{vi} \end{aligned} \tag{2.11}$$

where ϕ_{0v} is the cluster-specific random intercept. Equation 2.4 can be similarly reexpressed.

Once given details of each of the model components used in this document, the next step is to specify the abovementioned three-step strategy for the estimation of the unknown model parameters, as well as for performing variable selection and model choice, and to make the computation of the confidence intervals.

2.2 Three-step estimation strategy

As mentioned earlier, given the high dimensionality and complexity of the models specified in Equations 2.3, 2.4, and 2.11, we implement a three-step methodology consisting of the next proceedures. First, we perform the estimation, variable selection, and model choice via the boosting algorithm (Friedman, 2001; Hofner et al., 2014; Hothorn et al., 2020). We thereupon apply complementary pairs stability selection with per-family error rate (PFER) control to avoid falsely selecting covariates (Meinshausen & Bühlmann, 2010; Shah & Samworth, 2013). Lastly, we calculate confidence intervals for the relevant variables (Hofner et al., 2014). Specifications on this three-step strategy are offered directly after this paragraph in the subsequent lines.

2.2.1 Functional gradient descent boosting

In the context of structured additive models, formulations such as the expressed in Equations 2.3, 2.4, and 2.11 contain a large number of potential covariates linked to many complex alternative effects, and in consequence, the number of unkown parameters to estimate tend to be very large.

In this setting, three key methodological challenges arise. First, it is required an estimation method for the model, however classical alternatives perform poorer and poorer as data dimensionality and complexity increase (Johnstone & Titterington, 2009). Second, given the large number of possible covariates, it is also needed to identify a low-dimensional subset of covariates from the full data space containing all and only the relevant variables (*i.e.* variable selection), which is "almost" impossible or computationally burdensome for classical methods (Fenske et al., 2011). Finally, variables in additive regression models generally have different competing modeling possibilities (linear, nonlinear, and/or interaction effects), and thus, the challenge is not only to perform variable selection but also to choose the most appropiate functional form describing the relationship of each of the relevant covariates with the response variable (model choice).

One of the alternative estimation proceedures to overcome these challenges in structured additive models is to apply the functional gradient descent boosting algorithm to minimize the empirical risk (Bühlmann & Hothorn, 2007; Friedman, 2001). This algorithm is a regularization tech-

nique having the practical advantage of performing parameter estimation together with automatic variable selection and model choice (Bühlmann, 2006; Fahrmeir et al., 2013).

Algorithm

The boosting approach aims at minimizing the difference between the observed data and the model via the loss function:

$$\rho(y, \eta) \in \mathbb{R} \tag{2.12}$$

where $\eta = \beta_0 + \sum_{l=1}^{p} \beta_l w_l + \sum_{r=1}^{q} s_r(z_r) + s_{geo}(lon, lat) + s_{int_1}(z_d)w_g + s_{int_2}(z_e, z_f)$ (for the case of the model expressed in Equation 2.3, but similarly defined for the other cases) is a prediction function to be optimized. This loss function needs to be specified according to the model to be fitted. In the boosting approach, the goal is to iteratively solve the optimization of the expected loss function given by:

$$\hat{\eta} := \underset{\eta(\cdot)}{\operatorname{argmin}} \, E_{Y,W,Z}[\rho(y, \eta)] \tag{2.13}$$

where ρ is assumed to be differentiable and convex with respect to η (Schmid & Hothorn, 2008). Since $E_{Y,W,Z}[\rho(y, \eta)$ is in practice unknown, it is replaced by the empirical risk:

$$\sum_{i=1}^{n} \rho(y_i, \eta_i) \tag{2.14}$$

for observations $i = 1, ..., n$.

In η, let each of the unknown parameters and functions to estimate $\beta_0, ..., \beta_p$ and $s_1(z_1)$,..., $s_q(z_q), s_{geo}(lon, lat), s_{int_1}(z_d)w_g, s_{int_2}(z_e, z_f)$ represent a vector related to a specific block of covariates. These blocks are disjoint subsets of the data and are utilized as base-learnes, denoted as $\mathbf{b_0}$,..., $\mathbf{b_p}$ and $\mathbf{g_1}$,..., $\mathbf{g_q}, \mathbf{g_{geo}}, \mathbf{g_{int_1}}, \mathbf{g_{int_2}}$, respectively. These base-learners define the type of effect for each covariate, for instance in the parametric component of the model in Equation 2.3, $\mathbf{b_1}$ leads to a linear effect for variable w_1. In the case of effects for continuous variables (nonlinear, interaction, and spatial), every $\mathbf{g_1}$,..., $\mathbf{g_{int_2}}$ combines all the polynomials of the same covariate effect. Then, the algorithm is executed as specified by Friedman (2001) and Friedman et al. (2000):

1. Establish a maximum number of initial boosting iterations, m_{stop}. Then, initialize all the blocks $\beta_0, ..., \beta_p$ and $s_1(z_1), ..., s_{int_2}(z_e, z_f)$ with appropiate offset (starting) values $\beta_0^{[0]}, ..., \beta_p^{[0]}$ and $s_1^{[0]}(z_1), ..., s_{int_2}^{[0]}(z_e, z_f)$.

2. Set the iteration index $m = 1$ and compute the negative gradient of $\rho(y, \eta)$ evaluated at the previous iteration $\hat{\eta}_i^{[m-1]}$:

$$u_i^{[m]} = -\frac{\partial \rho(y_i, \eta_i)}{\partial \eta_i} \mid_{\eta_i = \hat{\eta}_i^{[m-1]}} \tag{2.15}$$

3. Obtain estimates for $\hat{\mathbf{b}}_{\mathbf{0}}^{[\mathbf{m}]}$,..., $\hat{\mathbf{b}}_{\mathbf{p}}^{[\mathbf{m}]}, \hat{\mathbf{g}}_{\mathbf{1}}^{[\mathbf{m}]}, ..., \hat{\mathbf{g}}_{\mathbf{int_2}}^{[\mathbf{m}]}$ by separately fitting each of the base-learners to the negative gradient (Equation 2.15). This process yields to obtain as many $u^{[m]}$ as the number of base-learners, *i.e.* as the number of covariate effects. Select the best-fitting base-learner in terms of minimization of the loss.

   - If the best-fitting base-learner is $\hat{\mathbf{b}}_{\mathbf{l}^*}^{[\mathbf{m}]}$, then update $\hat{\beta}_{\mathbf{l}^*}^{[\mathbf{m}]} = \hat{\beta}_{\mathbf{l}^*}^{[\mathbf{m-1}]} + \nu \hat{\mathbf{b}}_{\mathbf{l}^*}^{[\mathbf{m}]}$, for $\nu \in (0, 1]$, and leave the other blocks unchanged, *i.e.* $\hat{\beta}_{\mathbf{l}}^{[\mathbf{m}]} = \hat{\beta}_{\mathbf{l}}^{[\mathbf{m-1}]}$ for all $l \neq l^*$ and all the $\hat{\mathbf{s}}_{\mathbf{1}}^{[\mathbf{m}]}(\mathbf{z_1}), ..., \hat{\mathbf{s}}_{\mathbf{int_2}}^{[\mathbf{m}]}(\mathbf{z_e}, \mathbf{z_f})$ remain with the values of the previous iteration.
   - If the best-fitting base-learner is $\hat{\mathbf{g}}_{\mathbf{l}^*}^{[\mathbf{m}]}$, then update $\hat{\mathbf{s}}_{\mathbf{l}^*}^{[\mathbf{m}]}(\cdot) = \hat{\mathbf{s}}_{\mathbf{l}^*}^{[\mathbf{m-1}]} + \nu \hat{\mathbf{g}}_{\mathbf{l}^*}^{[\mathbf{m}]}$, for $\nu \in (0, 1]$, and leave the other blocks unchanged, *i.e.* $\hat{\mathbf{s}}_{\mathbf{l}}^{[\mathbf{m}]}(\cdot) = \hat{\mathbf{s}}_{\mathbf{l}}^{[\mathbf{m-1}]}(\cdot)$ for all $l \neq l^*$ and all the $\hat{\beta}_{\mathbf{0}}^{[\mathbf{m}]}, ..., \hat{\beta}_{\mathbf{p}}^{[\mathbf{m}]}$ remain with the values of the previuos iteration.

4. The algorithm is repeated until $m = m_{stop}$ by using the full set of base-learners again, including those obtained in the previous steps.

The entire implementation of the functional gradient descent boosting algorithm requires the specification of four tuning elements, namely the starting values (offset), the base-learners, the initial boosting iterations m_{stop}, and the parameter ν.

Tuning parameters

The choice of these tuning parameters in this research is as follows. Regarding the offset, in the models in this paper we decided to initialize the effect at

the mean as a starting value for the model intercept aiming at speeding up the algorithm's convergence (Fenske et al., 2011; Hothorn et al., 2020).

About the base-learners, in this document we use the simplest case, in which every block is related to only one covariate effect.

The most relevat tuning parameter for the algorithm is the number of boosting iterations (Friedman et al., 2000). To prevent overfitting, the optimal number of boosting iterations is chosen via cross-validated estimation of the empirical risk (Hothorn et al., 2020). By choosing the optimal number of iterations, the boosting algorithm also enables variable selection and model choice since only the most influential variables are picked with the appropriate functional form. By doing this, multicollinearity problems are avoided (Hofner et al., 2014).

The parameter ν, the step-lenght factor of the algorithm, has been found to be of relatively minor relevance for an appropiate execution of the boosting algorithm, nevertheless smaller values of ν increase the shrinkage and incidentally, the optimal number of boosting iterations becomes greater (Bühlmann & Hothorn, 2007; Schmid & Hothorn, 2008).

2.2.2 Stability selection

Once the model is fitted at the optimal number of iterations in step 1, we execute stability selection as proposed by Shah and Samworth (2013) to avoid the erroneous selection of non-relevant variables. By using subsampling procedures, this method simulates a finite number of random subsets of the data, and then, in each of these subsets, it controls the error rate for the number of falsely selected noise variables while selecting relevant variables in the fitting process of the boosting algorithm.

After this finite number of subsets have been fitted, the relative selection frequency per covariate effect is determined by calculating the proportion of subsets for which an effect is selected as relevant. All the effects with a relative frequency of selection equal or greater than a threshold previously specified are declared as stable effects. As a result of this selection, a parsimonious model is derived consisting exclusively of stable factors, in other words, we obtain a model with only non-zero regression coefficients. Regression coefficients for factors that are not selected as stable equal zero, indicating that they have no influence on the response variable. Setting these coefficients to zero is key, since it enables the variable selection and model choice processes.

In this thesis, we use 50 subsampling replicates and a threshold for the

relative selection frequency of 0.8, that is to say, for a covariate effect to be considered stable, it has to be selected as an influential predictor in at least 80% of the 50 random subsets. As shown in Meinshausen and Bühlmann (2010) results with a cutoff of between 0.6 and 0.9 do not significantly vary. Given the number of potential predictors and their alternative effects in our models, the cutoff of 0.8 corresponds to a PFER with a significance level of less than 0.05.

2.2.3 Pointwise bootstrap confidence intervals

Finally, 95% confidence intervals for the subset of effects selected as stable in step 2 are calculated by drawing 1000 random samples from the empirical distribution of the data using a bootstrap approach based on pointwise quantiles (Hofner et al., 2016). In this way, a stable effect is found significant if its corresponding 95% confidence interval does not contain zero.

3. Applications in development economics

"Some people hate the very name of statistics, but I find them full of beauty and interest....[T]heir power of dealing with complicated phenomena is extraordinary."
Francis Galton

In this chapter we discuss the appropriateness of utilizing structured additive regression models in development economics by applying this approach to examine three relevant phenomena in this field, namely use of time, violence against women, and poverty.

First, in Section 3.1 the goal is to comprehensively analyze the factors that explain the gap in time allocation to unpaid housework among partnered women and men in urban Mexico in 2020. In this study, we estimate a structured additive Gaussian model by using a data set composed of 16,167 observations and 30 theoretical covariates.

In Section 3.2 we study the risk factors for emotional IPV against women and girls in Mexico. To that end, we generate a data set with 35,004 observations and 39 covariates, to which we apply a structured additive probit model. The data set integrates ten information sources, allowing us to characterize IPV from a multilevel perspective, including the individual, re-

lationship, community, and societal levels. This section is based on Torres Munguía and Martínez-Zarzoso (2022).

Finally, in Section 3.3 we examine the effect of a set of potential risk factors on two of the lowest quantiles of income-to-poverty ratio distribution, namely the corresponding to poor and extremely poor families. Focusing on identifying heterogeneous effects according to the sex of the household head, we apply additive quantile models to a cross-sectional data set containing information on 4,434 women-headed and 14,877 men-headed households. For each model, we introduce 42 variables at the individual/household, community, and regional levels. The content of this section is based on Torres Munguía and Martínez-Zarzoso (2020, 2021).

Details about the background, theoretical framework, data, model, and results for each of these three studies are presented in the following sections of this chapter.

3.1 Understanding gendered inequalities in time allocation to unpaid housework among partnered women and men in Mexico

3.1.1 Background

Members of a family need to do the house chores, such as processing and preparing their meals, washing their clothes, and cleaning their house, to keep their home "livable" every day. The distribution of these activities is however uneven among household members, particularly when observing at the sex of the individuals (UN Women, 2019). Global reports overwhelmingly indicate that women disproportionately bear the burden of unpaid housework, spending around three times more time on these activities than men (UN Women, 2019).

Being time a limited resource, it is evident that the greater the amount of time a woman spends in unpaid housework the less time she can devote to income earning activities or leisure. As a result, women are excluded from engaging in the labor force, their social and economic empowerment is deteriorated, and/or a double-burden to employed women is implicitly imposed (Espino et al., 2020). The existence of these dissimilarities is particularly worrisome during and in the aftermath of the COVID-19 pandemic, given that the already existing gaps may be exacerbated as more people spend more time at home due to the emergency measures put in place to contain the virus propagation (Alon et al., 2021; ILO, 2020).

Examining how women and men differently allocate their time to unpaid housework is essential to understand one of the most alarming expressions of gender disparity in our society (UN Women, 2018, 2019). Prior research about the driving forces behind use of time patterns of men and women emphasizes the role played by factors such as education, income, family composition (number of children or elderly people in the household), and time devoted to other activities, namely leisure and paid work, on intrahousehold-decisions regarding time to unpaid domestic work (Begoña Álvarez, 2006; Bianchi et al., 2000; Datta Gupta & Stratton, 2010; Fang & McDaniel, 2017; Gimenez-Nadal & Molina, 2020). Nevertheless, the majority of research on time use concentrates on developed countries (United States or European

countries), and to the best of our knowledge, there is not a study analyzing the determinants of time allocated to unpaid domestic work for the case of Mexico.

Research on the matter is vital since their findings enable to obtain key insights that may help policy makers to identify strategies aiming to provide services, protection, or measures encouraging sharing of unpaid housework. The benefits of achieving this (reducing the women's domestic workload) are multiple. First and foremost this is a matter of equality and human rights, but it also generates better health, nutritional, educational, economic, and well-being conditions for the women and their families (IFAD, 2016). Furthermore, housework, seen as a transfer of services from the unpaid worker to others in an economy, is estimated to represent between 10 and 39 percent of the Gross Domestic Product, surpassing even the value from manufacturing and commerce sectors (UN, 2016).

Aiming to contribute to the debate on this matter, the goal of this research is to identify the factors explaining the gap in time allocation to unpaid housework among partnered women and men in urban Mexico in 2020. In order to achieve this, we apply a Gaussian model to a data set with more than 16,100 observations and 30 theoretical covariates. The recording unit of our data set is the woman, which enables us to describe not only their individual features, but also their intimate partners, peers, and families. To create these data, we use the information from the 2020 National Survey of Household Income and Expenditure (ENIGH). The inclusion of covariates from different thematic aspects enables us to capture diverse economic, social, and demographic aspects, such as income, education, peer networks, and household composition.

3.1.2 Research questions of this study

As previously stated, the overall objective of this research is to contribute to enhance our understanding on the factors behind the difference in time allocated to unpaid housework among partnered women and men in urban Mexico by using a structured additive Gaussian model. Specifically, we aim at producing empirical evidence to answer the following questions:

- Which factors are relevant to explain the gender gap in time to unpaid housework?

- How can the women's time allocation trade-off (time to paid work vs time to unpaid work vs time to leisure) be described? How does the time allocation trade-off of their partners affects women's time to unpaid housework?
- Is women's time to unpaid housework associated with age?
- What is the effect of the woman's economic situation on her time to unpaid housework?
- Do peer networks have a significant role in women's time distribution?
- Is time to unpaid housework linked to a particular type of families?
- Is the women's situation (regarding use of time, income, and age) relative to their men partners influencing the time the women devote to unpaid housework?

To provide elements to answer these questions, the analysis in this Section 3.1 is structured as follows. First, we present the conceptual framework used to analyze the time to unpaid housework. Then, we review the data and empirical strategy applied. Posteriorly, we present the results. Posteriorly we discuss these results in greater depth in order to provide some likely explanations according to previous findings. Finally, we present the conclusions.

3.1.3 Theory on the causes of and risk factors for gendered inequalities in time use

Based on previous research, we can identify disparate individual, relationship, and household factors found to be influencing time use distribution, and thus, the time allocated to housework.

With regard to the individual characteristics of the women, time spent in housework is first and foremost a trade-off against time to income earning activities and time to leisure (Rubiano Matulevich & Viollaz, 2019). Moreover, there is evidence suggesting that the time devoted to domestic activities is associated with age, economic situation, access to resources, ethnic origin, health condition, and education level (Gimenez-Nadal & Molina, 2020). In Kolpashnikova and Koike (2021), the authors found that, in Japan, Taiwan,

and the United States, women with low education levels tend to devote more time than women with high levels of education. Similar conclusions were achieved by Kan and Laurie (2016) analyzing data from United Kingdom and by Espino et al. (2020) for the case of Guatemala. Other studies have found that unemployment, low income, and older age groups do significantly more housework (Bianchi et al., 2000). Moreover, in a study with data from Norway, Anderssen and Wold (1992) also posited peer networks as a factor influencing the time a person devotes to paid and unpaid working activities, and to leisure. In particular, their results indicate that having friends supporting leisure-time physical activities significantly increases individuals reported time to these activities.

Existing research also indicates that woman's closest social circles (intimate partner, and family) have an influence on time to housework. About the intimate relationship, studies have found that women's time to unpaid housework decreases when the partner increases his housework, decreases his time to paid work or to leisure (Gimenez-Nadal & Molina, 2020). Other relevant factors found in literature are partner's age, his education level, and social connectedness (Álvarez & Miles, 2004; Fang & McDaniel, 2017; Gimenez-Nadal & Molina, 2020).

Finally, the family conditions also shape the distribution of time. Studies on the matter have pointed to the household structure, the number of members receiving income, the presence of children and elderly, and the kinship relationship among members as relevant factors for the allocation of time to domestic activities (Espino et al., 2020; Samtleben & Müller, 2022).

Interactions of variables have been suggested by different studies indicating that the relative power among the household members helps to explain the time distribution within the families (Gupta, 2006). This way, the individuals with lower power (in terms of income, education, or age) within the household or relationship are those devoting more time to unpaid domestic work (Bianchi et al., 2000; Datta Gupta & Stratton, 2010).

3.1.4 Data

Sources

After reviewing existing research on the matter, we identify the theoretical correlates associated with time use and then we map the official sources to obtain this information for the case of Mexico. The data used in this study

come from the 2020 ENIGH (INEGI, 2020). The ENIGH is a nationally representative household survey conducted every two years with the goal of generating official data about the income and expenditures of the Mexican families, and in its section IX, the survey inquires about use of time of people aged 12 years and over. From the total of respondents, for this research we exclusively consider the subset of heterosexual partnered women and men living together, being married, or in cohabitation in urban communities. In this subset of households, at least one of the partners carries out a paid working activity. We also excluded from the analysis in this research people with an indigenous ethnic origin. Time use of non-heterosexual households and indigenous is intended to be studied in further research by applying specific models to these two population groups.

Dependent variable

Statistical information on time to housework is extracted via self-reported responses to a question in the 2020 ENIGH (INEGI, 2020), asking surveyed people aged 12 years and over about time spent during the last week on different activities, including unpaid domestic work (washing, cleaning, cooking, ironing, sweeping, moping, etc.).

To create our response variable, gap in weekly hours allocated to unpaid housework between women and men, we calculate the arithmetic difference between the woman's weekly hours allocated to housework and her partner's weekly hours allocated to these tasks. This way, we capture how unequal is the time distribution between them. Positive values indicate that the woman is devoting more time to domestic work than her man partner, while negative values specify the opposite time distribution situation. Summary statistics are shown in the Table 3.1.

Table 3.1 Summary statistics of the gap in weekly hours allocated to unpaid housework between women and men

	Mean	**SD**	**Median**	**Min**	**Max**
-Gap in weekly hours allocated to unpaid housework between women and men	19.67	17.75	18.00	−64.00	99.00

Independent variables

The set of theoretically associated factors is initially proposed based on previous research findings and comprises features of the woman, her parnter,

and their families. In total, we identify 30 independent variables considered as relevant for time to unpaid domestic work. Of the total of independent variables, 10 describe characteristics of the woman, eight are related to their partners, and 12 variables aim to characterize their households. Definitions and sources for each covariate can be found in the Supplementary information 5.2.

See Tables 3.2 and 3.3 for the full list of independent variables included in this research.

Table 3.2 Summary statistics of continuous covariates in the model

Variable	Mean	SD	Median	Min	Max
Individual characteristics of the women					
-Woman's weekly paid working hours	18.22	22.69	3.00	0.00	120.00
-Woman's weekly leisure hours	18.56	15.28	14.00	0.00	99.98
-Woman's age	43.22	12.41	43.00	15	90
-Woman's income-to-poverty ratio	1.01	2.05	0.07	0.00	52.70
Relationship characteristics					
-Partner's weekly paid working hours	41.53	22.63	48.00	0.00	168.00
-Partner's weekly leisure hours	19.34	16.13	14.00	0.00	99.98
-Partner's age	46.00	12.81	46.00	16	91
-Partner's income-to-poverty ratio	2.81	5.48	2.01	0.00	276.53
-Share of woman's labor income in total couple's labor income	0.24	0.31	0.04	0.00	1.00
Household characteristics					
-Household members with income	0.64	0.25	0.67	0.11	1.00
-Children household members	0.18	0.20	0.17	0.00	0.75
-Senior household members	0.05	0.18	0.00	0.00	1.00
-Number of children	1.67	1.17	2.00	0	9

See Supplementary information 5.2 for definitions of independent variables.

Table 3.3 Summary statistics of categorical covariates in the model

Variable	Categories	N	%
Individual characteristics of the women			
-Educational lag	yes*	2216	13.7
	no	13951	86.3
-Access to social security	yes*	5513	34.1
	no	10654	65.9
-Education level	low*	8594	53.2
	medium	4388	27.1
	high	3185	19.7
-Disability	yes*	2431	15.0
	no	13736	85.0
-Social networks for care	low*	3840	23.8
	medium	2108	13.0
	high	10219	63.2
-Social networks for entrepreneurship	low*	12581	77.8
	medium	354	2.2
	high	3232	20.0
Relationship characteristics			
-Partner's education level	low*	8404	52.0
	medium	4026	24.9
	high	3737	23.1
-Partner's social networks for care	low*	3229	20.0
	medium	1991	12.3
	high	10947	67.7
-Partner's social networks for entrepreneurship	low*	12066	74.6
	medium	382	2.4
	high	3719	23.0
Household characteristics			
-Credit card	yes*	5892	36.4
	no	10275	63.6
-Access to food	yes*	2508	15.5
	no	13659	84.5
-Access to health services	yes*	2709	16.8
	no	13458	83.2
-Dwelling with adequate quality	yes*	858	5.3
	no	15309	94.7
-Access to basic housing services	yes*	942	5.8
	no	15225	94.2
-Type of household	nuclear*	12926	80.0
	extended	3105	19.2
	other	136	0.8
-Children at home	yes*	13159	81.4
	no	3008	18.6
-Parents at home	yes*	386	2.4
	no	15781	97.6

Reference categories are denoted with *.
See Supplementary information 5.2 for definitions of independent variables.

After having the variables of interest, we checked for plausibility and removed missing cases to render the data ready for analysis (see Suplementary information 5.3). The final data set is composed of 16,167 observations, corresponding to non-indigenous partnered women who, at the time of being

surveyed, were aged 12 or over, were married or cohabitating with a man partner, and living in an urban community in Mexico. This data set is freely available from Figshare at https://doi.org/10.6084/m9.figshare.21183271.

3.1.5 Model specification

We apply a structured additive Gaussian model to examine how and to what extent the set of factors is associated with the time gap to housework devoted by the woman and her partner.

Formally expressing our model, consider the *Gaussian* distributed variable y_i, denote the time gap in unpaid housework between the woman i and her partner, for $i = 1, .., 16167$ surveyed women. Consider also the vectors $w_1, ..., w_{17}$, of 17 categorical covariates, and $z_1, ..., z_{13}$, of 13 continuous independent variables (see Tables 3.2 and 3.3). Thus, the model is denoted:

$$y_i = \beta_0 + \sum_{l=1}^{15} \beta_l w_{il} + \sum_{r=1}^{6} s_r(z_{ir}) + \sum_{d=1}^{2} s_{int_d}(varying_{id}) + \sum_{e=1}^{4} s_{int_e}(surface_{ie}) + \varepsilon_i \tag{3.1}$$

where the terms at the right-hand-side in Equation 3.1 are:

- β_0 is the intercept of the model;
- $\beta_1, ..., \beta_{15}$ are the parameters measuring the effect of the 15 categorical covariates in the model (see Section 2.1.1);
- $s_1(z_1), ..., s_6(z_6)$ are functions in the model to capture potentially non-linear effects of the continuous covariates (see Section 2.1.2);
- $s_{int_1}(varying_1)$,...,$s_{int_4}(varying_4)$ are the model components for the following interactions between a continuous and a categorical variables (see Section 2.1.4):
  - woman's age by education level,
  - partner's age by education level,
  - woman's age by presence of children at home, and

  - woman's age by condition of cohabitation with her parents (or parents in-law).

- the functions $s_{int_1}(surface_1), ..., s_{int_4}(surface_4)$ represent the effect of the interaction between pairs of continuous covariates (see Section 2.1.4):
  - woman's weekly paid working hours by partner's weekly paid working hours,
  - woman's weekly leisure hours by by partner's weekly leisure hours,
  - woman's age by partner's age, and
  - woman's income-to-poverty ratio by partner's income-to-poverty ratio.

- $\varepsilon_1, ..., \varepsilon_{16167}$ are the error terms.

In sum, the regression model in Equation 3.1 is composed by 33 competing effects for 30 potentially associated factors (see Tables 3.2 and 3.3). Table 3.4 lists all the alternative effects associated to the 30 independent variables included in the full model.

Table 3.4 List of alternative effects by covariate in the full model

Variable	Alternative effects
Individual characteristics of the women	
-Woman's age in years	Linear and/or nonlinear
-Credit card	Linear
-Educational lag	Linear
-Access to social security	Linear
-Education level	Linear
-Woman's age by education level	Interaction
-Disability condition	Linear
-Social networks for care	Linear
-Social networks for entrepreneurship	Linear
Relationship characteristics	
-Woman's weekly paid working hours by partner's weekly paid working hours	Interaction
-Woman's weekly leisure hours by partner's weekly leisure hours	Interaction
-Woman's age in years by partner's age in years	Interaction
-Woman's income-to-poverty ratio by partner's income-to-poverty ratio	Interaction
-Share of woman's labor income in total couple's labor income	Linear and/or nonlinear
-Partner's education level	Linear
-Partner's age by education level	Interaction
-Partner's social networks for care	Linear
-Partner's social networks for entrepreneurship	Linear
Household characteristics	
-Share of household members with income	Linear and/or nonlinear
-Share of children household members	Linear and/or nonlinear
-Share of senior household members	Linear and/or nonlinear
-Number of children	Linear and/or nonlinear
-Access to food	Linear
-Access to health services	Linear
-Dwelling with adequate quality and sufficient space	Linear
-Access to basic housing services	Linear
-Type of household	Linear
-Woman's age by the presence of children at home	Interaction
-Woman's age by condition of cohabitation with her parents (or parents in-law)	Interaction

As can be seen in Table 3.4, the full model contains three alternative effects for the 30 independent variables. For categorical variables we introduce parametric effects (see Section 2.1.1). For continuous covariates we consider both linear and nonlinear effects as competing alternatives (2.1.2). Finally, we consider interactions of variables. On the one hand, we include interactions between a continuous and a categorical variable to capture possible disparities in the effect of age by different levels of education, and according to the family composition (presence of children and/or parents). On the other hand, interactions of continuous variables aim to estimate the potential effects regarding the situation of the woman relative to her partner (about age, income, and time use).

The resulting model expressed in Equation 3.1 has high dimensionality

due to the large number of covariates considered and its estimation cannot be computed by traditional models. Consequently, we proceed with the three-step strategy described in Section 2.2. Details of their implementation are described in the following lines and the results are presented in Section 3.1.6.

Implementation details

The first step of the methodology, the boosting algorithm (see Section 2.2.1), is applied with 2000 initial iterations, and posteriorly, to prevent overfitting and to determine the optimal number of iterations, we performed cross-validation. The model is optimized at 695 iterations.

The stability selection, as explained in 2.2.2, is used to avoid the erroneous selection of non-relevant variables. Specifically for this research, considering the number of potential effects associated to the total number of variables, we set a cutoff of 0.8 and 50 complementary pairs for the error bounds. This configuration corresponds to a PFER with a significance level of 0.0316.

Lastly, we calculate 95% confidence intervals for the subset of stable effects from step 2. To do this, we draw 1000 random samples from the empirical distribution of the data using a bootstrap approach based on pointwise quantiles (see Section 2.2.3).

All computations are implemented in the R package "mboost" (Hothorn et al., 2020). The corresponding code to replicate these results can be found in the Supplementary information 5.4 and is also freely available from Figshare at https://doi.org/10.6084/m9.figshare.21183271.

3.1.6 Application results

From the total of 30 potential covariates linked to 33 alternative effects, after applying the three-step methodology, only six effects corresponding to seven covariates are selected as relevant. Table 3.5 shows the list of selected covariates having a significant effect on the gap of time to housework. Given that our response variable measures how uneven is the weekly time allocation to unpaid housework between partnered women and men, with positive values indicating that the woman devotes more time to housework than her man partner, coefficients can be interpreted as the estimated change in hours per week associated with a particular variable. This way, for categorical covariates, parameters indicate the difference in the estimated coefficient for a category with respect to the reference category coefficient. For continu-

ous covariates with linear effects, the parameter indicates the change in the time gap, expressed in weekly hours, per unit change in the continuous independent variable. For continuous covariates with nonlinear effects or with interacting effects, interpretation is best done by visualizing the corresponding figures.

Table 3.5 Selected variables associated with gap in weekly hours allocated to unpaid housework between women and men

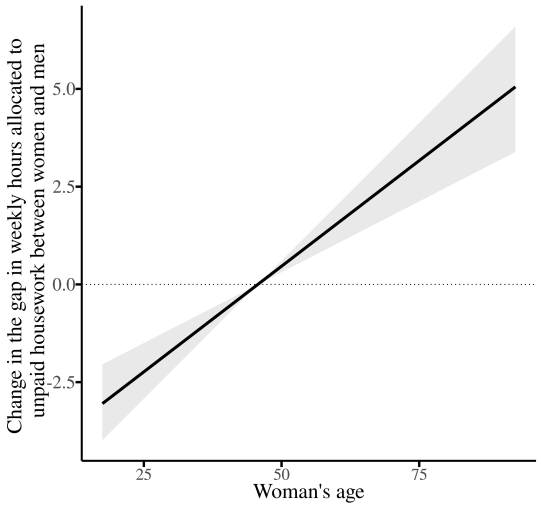
Variable	Categories	Coefficient [95% CI]
Individual characteristics of the women		
-Woman's age		Linear, slope: 0.108 (Fig. 3.1)
-Woman's weekly paid working hours		Nonlinear (Fig. 3.2)
-Education level	low*	
	medium	
	high	-2.492 [-3.08, -1.95]
Relationship characteristics		
-Woman's weekly leisure hours by partner's weekly leisure hours		Interaction surface (Fig. 3.3)
-Share of woman's labor income in total couple's labor income		Nonlinear (Fig. 3.4)
Household characteristics		
-Number of children		Linear, slope: 1.283 (Fig. 3.5)

Reference categories are denoted with *.
Empty cells indicate that the corresponding effect is not stable and therefore it is set to zero.

Regarding the individual characteristics of the women, it is found that her age, the time spent in paid work, and her education level are significantly associated with the gap in time to unpaid domestic work.

About women's age, it is observed in Fig. 3.1 that a more equal time distribution between women and men is observed among partnered young women, regardless of partner's age. Differences in time to unpaid housework grows with woman's age at a rate of 0.108 hours for every additional year of age, which is equivalent to a difference of almost seven weekly hours between a woman aged around 20 years old and one aged about 90 years old.

Fig. 3.1 Linear effects of women's age on the gap in weekly hours allocated to unpaid housework between women and men



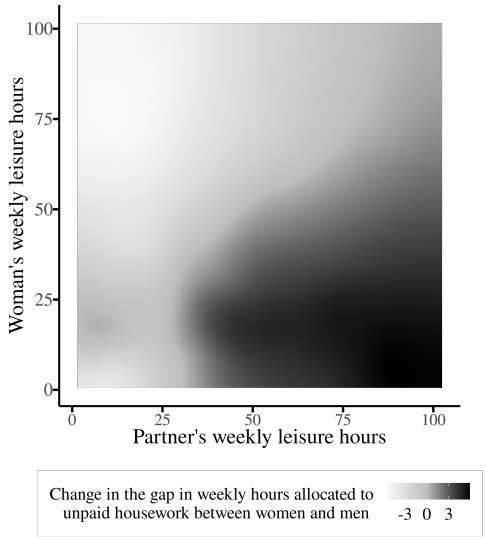
Women's weekly hours devoted to paid work is observed to have a non-linear decreasing effect on the gap of time to housework, which indicates that the difference between the woman's and man's time to domestic work decreases as the woman increases her time to paid work (Fig. 3.2). It is important to highlight that at the rightest side of the plot, the effect of increasing the paid working hours on the time gap to unpaid housework is constant.

Fig. 3.2 Nonlinear effects of women's weekly paid working hours on the gap in weekly hours allocated to unpaid housework between women and men

Women's education level is also significantly associated with time gap to unpaid housework. The results indicate that women with a high level of education (at least a completed university degree) tend to have a better time distribution with their partners, in comparison to women having a low level of education. The difference is estimated to be between 1.95 and 3.1 weekly hours (see Table 3.5).

About the features of the relationship, *i.e.* the woman's situation relative to her partner, after applying our methodology, we find that variables woman's weekly leisure hours interacting with partner's weekly leisure hours and share of woman's labor income in total couple's labor income have a relevant effect on time gap to unpaid housework. On the interaction effect of weekly leisure hours with partner's weekly leisure hours (Fig. 3.3), we observe that the already existing gap against women's time to housework tends to be larger at the bottom right corner of the plot, which, generally speaking, corresponds to few leisure hours for the women (even no time for leisure) but many hours of leisure for the men (more than approximately 30 hours).

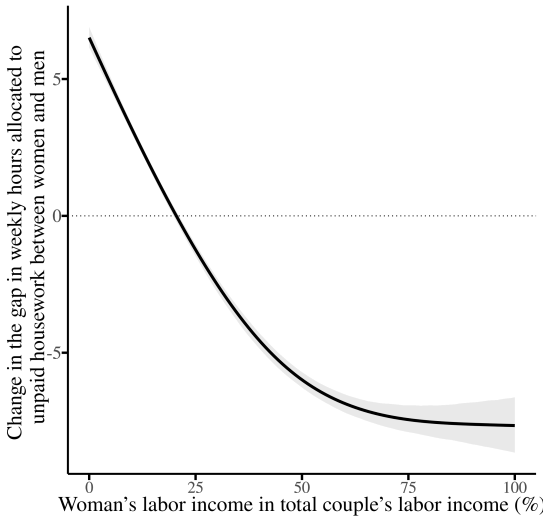
Fig. 3.3 Interaction effects of weekly leisure hours with partner's weekly leisure hours on the gap in weekly hours allocated to unpaid housework between women and men



The darker the color the larger the gap in time to housework between women and men

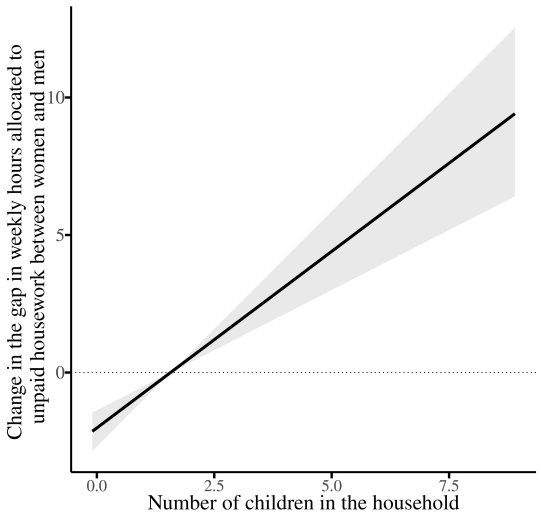
Furthermore, it is also observed a significant relationship of the contribution of woman's labor income in total couple's labor income and gap in time to housework (Fig. 3.4). On the whole, a better distribution of unpaid housework time is found in couples in which the woman has a greater contribution to total labor income. Inequality in time to domestic work increases as the difference between woman's and man's labor income widens. In fact, the time gap increases in up to 6.5 weekly hours when all the income of the couple is provided by the man (see the upper right corner of the Fig. 3.4).

Fig. 3.4 Nonlinear effects of share of woman's labor income in total couple's labor income on the gap in weekly hours allocated to unpaid housework between women and men



About the household characteristics, we found that the composition of the family is found to have a significant association with the time gap to housework. In particular, it is expected that families having children exhibit a worse time distribution to housework against women. As the number of children members increases in the household, the time gap to unpaid domestic work also raises by 1.28 weekly hours per child (Fig. 3.5). This way, women living in families without children have a time gap to housework lower in about four hours in contrast to women in households with three children.

Fig. 3.5 Linear effects of number of children in the household on the gap in weekly hours allocated to unpaid housework between women and men



3.1.7 Discussion of the application results

It is important to keep in mind that all the factors examined in the previous section indicate statistical relationships between the selected variables and the time gap to unpaid housework, and even though they do not necessarily imply causality, they provide evidence about crucial aspects for the studies on time use among partnered women in urban Mexico, and some potential explanations can be derived based on existing studies and theories.

As previous studies on time use have found, time allocation to housework is first and foremost a trade-off against time to income earning activities and time to leisure (Rubiano Matulevich & Viollaz, 2019). Our results about women's time to paid work and to leisure corroborate this: a person cannot increase their housework time without decreasing time devoted to other activities.

Specifically, time gap to housework is also associated with the partner's time to leisure, but it was found to be independent from the partner's time to paid work. This result is particularly interesting since it suggests the existence of an asymmetric role-related time use allocation to housework. Previous studies indicate that these inequalities are the result of stereotypes

and discriminatory social norms that put women at a disadvantage in the process of making intra-households decisions, and thus, the more time the man spends on leisure, the more the woman dedicates to housework (Álvarez & Miles, 2004).

Woman's age matter in determining the amount of time devoted to unpaid housework, which might be suggesting the existence of a generational improvement on the gender roles observed in the time distribution: traditional roles behind the uneven distribution of time are mainly associated with elder women, while the younger ones have a less unequal time distribution (Nitsche & Grunow, 2016).

A key factor found to be significantly associated with time gap is labor income. Results lend support to previous theories and studies highlighting the importance of women's own income in housework time, and how the larger the economic dependence of women relative to their men partners, the greatest the gap against women's time to housework (Gupta, 2006; Kan & Laurie, 2016).

Women's education level is also a variable constantly found as relevant for time distribution and our results are in line with the commonsense idea indicating that women with a higher level of education have a better distribution of time to housework with their partners than women with lower educational achievements (Kolpashnikova & Koike, 2021). Although more detailed statistical information is required to inquire into the reasons for this result, a potential explanation could be that having achieved a high level of education promotes pro-gender equality attitudes and behaviors, which is reflected in a more equal distribution of time to housework with their partners.

Another important insight comes from the result about the effect of the share of children household members. This agrees with previous studies for multiple countries (Giurge et al., 2021), which indicate that women unevenly shoulder the major burden of childcare and their associated responsibilities.

3.2 Emotional IPV against women and girls with children in Mexican households

3.2.1 Background

According to the 2016 National Survey on the Dynamics of Household Relationships (ENDIREH), the most prevalent act of IPV against women in Mexico is emotional abuse, affecting approximately 40.1% (17.4 million) of all ever-partnered women aged 15 years or over (INEGI, 2016c).

The impact of emotional IPV is severe. Half of these victims experienced stress, depression, insomnia, and loss of appetite, and about 8.6% (1.5 million) of them have thought about killing themselves or have already attempted suicide (INEGI, 2016c). IPV also affects other family members, especially the children (Sturge-Apple et al., 2012), placing women with children at a particular risk regarding IPV (Peek-Asa et al., 2017).

Given the above-mentioned consequences, gaining a better understanding of what drives the risk of emotional IPV victimization is of paramount importance. One of the most widely used approaches to study the multifaceted nature of violence, including IPV, is the ecological model. According to this approach, violence can be explained as a result of the interaction and convergence of multiple factors at four interrelated levels: individual, relationship, community, and society (WHO, 2012).

For the Mexican case, findings from studies using the ecological model suggest that being young (Castro et al., 2006; Villarreal, 2007), having a low education level (Avila-Burgos et al., 2009; Jaen Cortés et al., 2015; Rivera-Rivera et al., 2004), a low socioeconomic status (Castro & Casique, 2008; Castro et al., 2006), and/or being in a relationship with a young man (Casique & Castro, 2014) who abuses alcohol/drugs (Mojarro-Iñiguez et al., 2014), displays controlling behavior (Frías, 2017), has a history of violence victimization/perpetration (López Rosales et al., 2013), or who is unemployed (Valdez-Santiago et al., 2013) are risk factors for IPV victimization.

Despite the importance of these findings, there are still some factors from the ecological approach, whose relevance is generally acknowledged in international studies (UNiTE Working Group, 2019; WHO, 2012), but their effects have not been examined in-depth -or not at all- for Mexico. Three of them belong to the individual level: young age at first childbirth, unwanted sexual initiation at an early age, and lack of a pro-gender equality attitude.

Four additional understudied factors belong to the relationship level: getting married young and not by choice, a lack of decision-making power, unequal distribution of housework to women's detriment, and a lack of peer networks. At the community level, there are two factors: risks of women living in communities with unequal income distribution and a low level of women's participation in the public sphere. Similarly, at the societal level, the effects of living in a society with low quality of government, high corruption levels, and high rates of criminal activity on IPV victimization have also not yet been examined. The analysis of these factors is vital since identifying their effects could help policy-makers not only to design specific strategies to protect current victims but to develop early interventions targeted at high-risk communities and regions, and the most vulnerable population groups, in order to prevent future aggressions (Heise, 2011).

Intending to contribute to the discussion on the emotional IPV risk factors in Mexico, emphasizing the above-mentioned understudied factors, we examine how and to what extent a set of theoretical factors are linked to women's probability of victimization. To that end, we apply a probit model to data on the victimization experiences of 35,004 women and girls with children in Mexico. The population studied includes women and girls aged 15 years and over, which allows us to capture how the risks vary over their lifetime, from adolescence to old age, while controlling for the rest of the factors in the model. Our data contain more than 40 potential variables at the four levels of the ecological model, taken from ten official data sources. The creation of this data set enables us to overcome two shortcomings in IPV studies identified by Krug et al. (2002) in the *World report on violence and health*: the inclusion of a limited number of potential risk factors and the lack of characterization of the community and society where violence occurs.

3.2.2 Research questions of this study

The overall goal of this study is to contribute to improving the knowledge of the risk factors for emotional IPV against women and girls in Mexico by using a structured additive model. In particular, we aim at providing comprehensive and evidence-based answers to the following key aspects:

- Which independent variables at the individual, relationship, community, and societal levels are relevant to explain emotional IPV?

- How do victimization risks vary across the life course? Does this effect differ across demographic groups (indigenous origin or by education level)?
- Do continuous covariates (such as women's age, women's age at first childbirth, or women's age at her first sexual intercourse) have a linear or a nonlinear effect on victimization risks?
- Is the situation of the women (regarding age and income) relative to their men partners altering their risks of emotional IPV victimization?
- Are previous women experiences (such as age at and consent to marriage or cohabitation, and age at and consent to first sexual intercourse) associated with their current IPV victimization risks?
- Do gender-related issues (such as housework distribution, decision-making power, and pro-gender equality attitudes) have an effect on emotional IPV victimization?
- How can we describe the economic, demographic, and political features of the communities and societies where the emotional IPV victimization most frequently occurs? Do inequality, criminality, corruption, and quality of public services play a significant role?

To answer these questions, this Section 3.2 is divided as follows. First, we briefly introduce the theoretical framework to understand IPV victimization. Posteriorly, the data and model specification are presented. Then, the findings from the empirical methodology are commented. After this, a discussion highlighting some potential explanations for the results is provided. Finally, we present the conclusions of this research.

3.2.3 Theory on the causes of and risk factors for IPV

One of the most widely used approaches to study the multifaceted nature of violence, including IPV, is the ecological model. According to this model, IPV is grounded in a combination of factors operating at four different levels: individual, relationship, community, and societal. It is critical to note that each of these factors interacts not only with the rest of the factors within their corresponding level but also with those from the other levels. These interactions play a crucial role since no single factor can explain IPV, but

rather a myriad of them shape the women's victimization risks (WHO, 2012). To briefly discuss the factors identified as relevant across studies from different countries most consistently, we present some findings at each level of the ecological model to analyze the Mexican case.

Individual-level factors

At the individual level there is evidence suggesting that IPV is more prevalent among women and girls from minority groups, with a low level of economic empowerment, who had their first childbirth at an early age, their first sexual intercourse at a young age and/or against their wishes, and who lacks a pro-gender equality attitude (WHO, 2012). Findings supporting the relevance of these variables can be found in the study by Oduro et al. (2015) for the cases of Ecuador and Ghana, by Stöckl et al. (2014) using data from a multi-country study, by Cameron and Tedds (2021) with data from Canada, and by Caetano et al. (2005) with data from the United States.

Woman's age has also been found to be significant for IPV, yet findings suggest that this association has a context-specific effect. Using data from the United States, Walton-Moss et al. (2005) found that larger risks are observed among young women. By contrast, Wilson (2019), using data from 36 countries, concluded that particularly for emotional and physical IPV, the association with age is described by an inverted U-shaped curve.

Findings also indicate that the intersection of women's age with other demographic characteristics alters the IPV risks. For instance, Heidinger (2021) found that the gap in victimization risks between indigenous and non-indigenous Canadian women varies over their lifetime, reaching a maximum between 25 and 44 years old. A similar pattern is described for the interaction of age with education level in the report by Oficina de Violencia Doméstica (2021) with data from Argentina.

Relationship-level factors

The second ecological level captures the features of the woman's closest social circles: her intimate relationship and her relationship with peers and family.

Regarding the intimate relationship, some studies have found a number of the partner's characteristics correlated with IPV: young age, low education levels, frequent alcohol consumption, and living in economic disadvantage (National Center for Injury Prevention and Control, 2020; WHO, 2012).

These results are confirmed by Stöckl et al. (2021) with data from Sub-Saharan Africa, Caetano et al. (2001) for the United States, Stöckl et al. (2012) for the German case, and Ahmadabadi et al. (2020) for Australia.

Nevertheless, the above-mentioned features are not risk factors *per se* but instead refer to the woman's situation relative to her partner. For instance, Rapp et al. (2012) concluded that lower IPV risks are observed among couples with the same education level in India and Bangladesh. In the same vein, results reported in Abramsky et al. (2019) for the case of Tanzania, and in Reichel (2017) with data from the European Union countries, lend support to the idea that discrepancies between the woman's economic status and that of her partner lead to higher IPV risks. Similarly, Chaurasia et al. (2021) found that a large age gap exacerbates the likelihood of experiencing IPV in India. Moreover, women's autonomy has been found to be negatively correlated with IPV in Pakistan (Mavisakalyan & Rammohan, 2021) and Turkey (Yilmaz, 2018).

Regarding the woman's relationship with her peers and family, research shows that an unequal distribution of housework, overcrowding, and inadequate social support networks are factors that increase IPV. Among others, such findings have been reported by Wright (2015) with data from Chicago, Nguyen et al. (2018) for pregnant women in Vietnam, and Plazaola-Castaño et al. (2008) in three autonomous communities in Spain.

Community-level factors

At the community level, findings indicate that women living in urban settlements, in communities with high crime incidence, high concentration of immigrants, unfavorable socioeconomic circumstances, and/or gender-inequitable conditions are at greater risk of IPV (UNiTE Working Group, 2019; WHO, 2012). Some papers coming to these conclusions are those by Dias et al. (2020) studying six European cities, Lauritsen and Schaum (2004) and Voith et al. (2021) with data from the United States, and Ackerson and Subramanian (2008) for India.

Societal-level factors

At the fourth level of the ecological model, as found by Gillanders and van der Werff (2020) for the case of African countries with data from the Afrobarometer, Gashaw et al. (2018) for Ethiopia, and González and Rodríguez-Planas

(2020) using survey data from 28 European countries, the most consistent risk factors at the societal level include low quality of government, high crime incidence, social instability, and high prevalence of sexist norms and beliefs.

Previous research analyzing IPV in Mexico

Even though studies for Mexico have tended to apply the ecological model, they have almost exclusively analyzed the association of individual- and relationship-level factors with IPV.

About the individual-level factors, Castro et al. (2006) and Villarreal (2007) used data from the 2003 ENDIREH to show that younger women are potentially more at risk of IPV. In addition to age, Jaen Cortés et al. (2015) and Rivera-Rivera et al. (2004) found that women's education level is negatively associated with IPV victimization. Moreover, IPV is also more prevalent in women from low socioeconomic backgrounds (Castro & Casique, 2008; Castro et al., 2006).

Regarding the relationship-level factors, Casique and Castro (2014) and Castro et al. (2006) found that women with a young partner are more likely to suffer from IPV. Moreover, Rivera-Rivera et al. (2004), Esquivel-Santoveña et al. (2020), and Terrazas-Carrillo and McWhirter (2015) showed that other key IPV risk factors are the partner's heavy drinking and controlling behavior. By examining data from Monterrey in Mexico, López Rosales et al. (2013) found that higher risks are expected in women whose partners have a history of violence perpetration and/or victimization. The partner's socioeconomic disadvantages (low education level or unemployment) are also expected to be risk factors for IPV (Alvarado-Zaldívar et al., 1998; Avila-Burgos et al., 2009; Valdez-Santiago et al., 2013).

The community level remains largely understudied for Mexico. Only Castro and Casique (2009) distinguished between IPV risks in urban and rural communities, while Valdez-Santiago et al. (2013) analyzed data from some indigenous regions in Mexico to study the prevalence and severity of IPV and introduced covariates such as community type and poverty level in the municipality.

Concerning the societal level, only a handful of papers have considered single factors at this level of analysis for Mexico. García-Ramos (2021), analyzing state-level data over time, found that divorce laws significantly affect IPV in the long term, while Sterling (2018) argued that a sexist culture is strongly linked to a high risk of IPV in Mexico.

3.2.4 Data

Sources

After identifying a set of theoretical factors at the four levels of the ecological model, we map the official data sources containing this information for Mexico.

Our main source of information is the 2016 ENDIREH, from which we obtain data at the individual and relationship levels. The ENDIREH is a nationally representative household survey conducted by Mexico's National Institute of Statistics and Geography (INEGI). This survey aims to produce information on the violence experienced by women and girls aged 15 years and over in Mexico. The survey explores four types of violent acts, namely physical, sexual, economic, and emotional, which occur in the contexts of the community, workplace, and school environments, in the family, and within intimate relationships. For this research, we only use information from the questionnaire referring to heterosexual married or cohabiting women.

To characterize the community and societal levels, we identify in the ENDIREH the municipality and state where the respondent lives. Then, we merge the information about the individual and relationship levels from the ENDIREH with the estimations from the official poverty data generated by the National Council for the Evaluation of Social Development Policy (CONEVAL), marginalization data from the National Population Council (CONAPO), the municipal geographical information and homicide records collected by the INEGI, the human development index produced by the United Nations Development Program (UNDP), information from the 2015 Intercensal Population Survey, the 2016 National Survey on Victimization and Perception of Public Safety (ENVIPE), the 2015 National Census of Municipal and Delegation Governments (CNGMD), and from the 2015 National Survey of Quality and Governmental Impact (ENCIG). More details on these sources can be found in CONAPO (2016), CONEVAL (2020), INEGI (2015a, 2015b, 2015c, 2016b, 2016c), and UNDP (2019).

Dependent variable

Information on emotional IPV victimization is produced via self-reported responses to a question in the ENDIREH asking about the occurrence of 15 acts suffered in the context of their current or previous relationship in the

preceding 12 months, i.e. between October 2015 and October 2016 (see Table 3.6 for the list of acts and behaviors included as emotional violence).

Table 3.6 Acts of emotional IPV captured by the 2016 ENDIREH

Type of act	**Behaviors included**
Social isolation	-Forbidding the woman to leave the house, locking her up, or stopping her from having visits. -Turning children or relatives against the woman.
Threats	-Threatening the woman about abandoning her, to harm her, to take the children or to kick her out the house. -Threatening the woman with a weapon. -Threatening the woman to kill her, to kill himself or to kill the children.
Humiliation	-Humiliating her, degrading her, comparing her with other women or calling her ugly. -Blaming her on cheating on him.
Indifference	-Ignoring her, embarrassing her, not taking her into account or not giving her affection. -Stop talking to the woman.
Intimidation and stalking	-Making her feel scare. -Stalking her, spying her, following her around, showing up suddenly in places. -Calling or texting the woman repeatedly to know her location, if she is with someone and what she is doing. -Destroying, throwing, or hiding personal or family property. -Monitoring woman's mails or cellphone and demanding passwords. -Reproaching and getting angry with the female because household chores are not done in the way the male partner wants, because food is not done or because he considers she does not fulfill her obligations.

Source: INEGI (2016c) .

Possible responses to the question about IPV victimization are "many times", "sometimes", "once" and "never". Given that the frequency associated with "many times" and "sometimes" is not precisely defined but rather is left to the respondent's own judgment, we decide to generate a binomial variable by dichotomizing the answers into "yes" or "no" where the first three responses are considered as "yes". This allows us to focus specifically on the probability of experiencing emotional IPV. Summary statistics for the response variable in this study are shown in Table 3.7.

Table 3.7 Summary statistics of the response variable

	Categories	**N**	**%**
-Victim of emotional IPV in the last 12 months	no	25624	73.2
	yes	9380	26.8

Independent variables

Following the ecological approach and previous studies, we map the data sources with information on the matter and identify in them the available theoretical factors proposed in the literature review (see Section 3.2.3).

In total, we identify 39 independent variables considered as risk factors under the ecological approach, namely six variables at the individual level, 13 at the relationship level, 14 corresponding to the community, and six describing the societal level.

The full list of potential explanatory variables included in this study, together with their summary statistics, is listed in Tables 3.8 and 3.9. Definitions and sources for each covariate can be found in the Supplementary information 5.5.

Table 3.8 Summary statistics of continuous covariates in the model

Variable	Mean	SD	Median	Min	Max
Individual-level covariates					
-Woman's age	40.59	14.13	38	15	80
-Woman's income	861.50	1610.77	0	0	6000
-Woman's age at first childbirth	20.16	3.56	20	11	30
-Woman's age at her first sexual intercourse	18.23	3.15	18	9	28
Relationship-level covariates					
-Woman's age at marriage or at cohabitation	20.01	3.92	19	10	33
-Partner's age	43.87	14.68	41	15	83
-Partner's income	3877.44	2942.41	4000	0	12 000
-Overcrowding	2.41	1.04	2	0.20	5.25
Community-level covariates					
-Women homicide rate	22.73	20.80	17.30	0	343.21
-Men homicide rate	194.03	191.85	141.36	0	2392.34
-Total homicide rate	106.84	102.46	77.28	0	1142.42
-Women's household headship	0.25	0.05	0.26	0.06	0.39
-Migration of women	0.03	0.03	0.03	0	0.26
-Migration of men	0.04	0.03	0.03	0	0.26
-Gini index	0.40	0.03	0.40	0.30	0.58
-Human development index	0.75	0.07	0.76	0.42	0.94
-Municipal functional capacities	0.36	0.18	0.33	0	0.86
-Women's economically active population	67.82	6.46	69.17	18.32	83.80
-Men's economically active population	31.82	10.32	34.36	2.51	52.25
-Women's political participation	0.22	0.11	0.21	0	0.88
Societal-level covariates					
-Common crimes against women	23 134.74	6198.48	21 544.71	12 388.75	40 653.29
-Common crimes against men	25 837.98	7543.29	23 039.32	16 477.31	51 554.99
-Dark figure of common crimes against women	92.11	2.82	92.23	87.70	98.06
-Dark figure of common crimes against men	92.37	2.11	92.29	87.61	96.86
-Corruption	0.87	0.05	0.89	0.75	0.95
-Satisfaction with public services	0.39	0.09	0.40	0.24	0.54

See Supplementary information 5.5 for definitions of independent variables.

Table 3.9 Summary statistics of categorical covariates in the model

Variable	Categories	N	%
Individual-level covariates			
-Indigenous origin of the woman	no*	24301	69.4
	yes	10703	30.6
-Formal education level of the woman	low*	12591	36.0
	medium	20160	57.6
	high	2253	6.4
Relationship-level covariates			
-Woman's consent to first sexual intercourse	yes*	34310	98.0
	no	694	2.0
-Woman's consent to marriage or cohabitation	no*	2197	6.3
	yes	32807	93.7
-Pro-gender equality attitude	low*	1698	4.9
	medium	15809	45.2
	high	17497	50.0
-Division of housework among household members	only women*	22692	64.8
	both	7417	21.2
	only men	4895	14.0
-Woman's level of autonomy within the relationship to make decisions about her sexual life	low*	1646	4.7
	medium	29986	85.7
	high	3372	9.6
-Woman's level of autonomy within the relationship to make decisions about her professional life and use of economic resources	low*	1630	4.7
	medium	11582	33.1
	high	21792	62.3
-Woman's level of autonomy within the relationship to make decisions about her participation in social and political activities	low*	1514	4.3
	medium	13645	39.0
	high	19845	56.7
-Social networks	low*	467	1.3
	medium	4217	12.0
	high	30320	86.6
-Level of social interaction reported by the woman	low*	7860	22.5
	medium	25074	71.6
	high	2067	5.9
Community-level covariates			
-Social marginalization	very low*	17930	51.2
	low	7031	20.1
	medium	4811	13.7
	high	4305	12.3
	very high	927	2.6
-Type of community	rural*	241	0.7
	low urban	2765	7.9
	medium urban	12205	34.9
	high urban	19793	56.5

Reference categories are denoted with *.
See Supplementary information 5.5 for definitions of independent variables.

After combining the abovementioned data from the different sources and levels into a single, unified data set (see Supplementary information 5.7 for a detailed description of this data integration process), we checked for plausibility, detected outliers, and removed missing cases to prepare the data for the analysis (a description of this data cleaning process can be found in

Supplementary information 5.8). The final data set is composed of 35,004 observations, which correspond to women who, at the time of being surveyed, were aged 15 or over, were married or cohabitating with a male partner, and had had at least one child. This data set is freely available from Figshare at https://doi.org/10.6084/m9.figshare.21183271.

3.2.5 Model specification

In this study we apply a structured additive probit model to deal with the dichotomous nature of our dependent variable, whose binary outcome indicates whether the woman surveyed has suffered from emotional IPV during the reference period.

To formally express this design, let the variable y_i, following a $Bernoulli(\mu_i)$ distribution with probability $\mu_i \in [0, 1]$, indicate whether or not the woman i suffered ($1 = TRUE$) from emotional IPV, for $i = 1, ..., 35004$ observations. Also consider the set of 13 categorical $w_1, ..., w_{13}$ and 26 continuous covariates $z_1, ..., z_{26}$ (see Tables 3.8 and 3.9). The binomial model is given by:

$$\begin{aligned} h(\mu_{vi}) = \beta_0 + \sum_{l=1}^{13} \beta_l w_{ivl} + \sum_{r=1}^{26} s_r(z_{ivr}) + s_{geo}(lon_{vi}, lat_{vi}) + \\ \sum_{d=1}^{4} s_{int_d}(varying_{ivd}) + \sum_{e=1}^{5} s_{int_e}(surface_{ive}) + \phi_{0i} + \varepsilon_{vi} \end{aligned} \tag{3.2}$$

where the link function $h(\mu)$ is the inverse standard normal distribution of the women's likelihood of emotional IPV victimization, modeled as an additive combination of their risk factors. Similar to Equation 2.4, the model components of Equation 3.2 are:

- β_0 is the constant term for the model intercept;
- $\beta_1, ..., \beta_{13}$ are the unknown regression parameters for the effect of the 13 categorical covariates (see Section 2.1.1);
- $s_1(z_1), ..., s_{26}(z_{26})$ are smooth functions for the nonlinear effects of the 26 continuous covariates (see Section 2.1.2);
- $s_{geo}(lon, lat)$ is a component to model spatial effects of the geographic coordinates lon and lat for each municipality's centroid (see Section 2.1.3);

- $s_{int_1}(varying_1), ..., s_{int_4}(varying_4)$ are the model components capturing four interactions between continuous and categorical variables (see Section 2.1.4):
  - age of the woman by indigenous origin,
  - age of the woman by education level,
  - age of the woman at her first sexual intercourse by condition of consent, and
  - age of the woman at marriage or at cohabitation by condition of consent.
- parameters $s_{int_1}(surface_1), ..., s_{int_5}(surface_5)$ denote the interaction effect between pairs of continuous covariates (see Section 2.1.4):
  - age of the woman by age at first childbirth,
  - age of the woman by age at her first sexual intercourse,
  - age of the woman by age at marriage or at cohabitation,
  - age of the woman by age of the husband or partner, and
  - woman's reported monthly earned income by reported husband's or partner's reported monthly earned income.
- ϕ_0 is the cluster-specific random intercept due to the hierarchical data structure, in which individual observations are connected to the information for the municipalities, and these, in turn to the state information (see Section 2.1.5); and,
- $\varepsilon_1, ..., \varepsilon_n$ are the error terms.

All in all, the structured additive probit model expressed in Equation 3.2 consists of 74 potential effects related to 39 theoretical covariates (see Tables 3.8 and 3.9) plus spatial and random effects. Table 3.10 presents the list of all these alternative effects included in the full model.

Table 3.10 List of alternative effects by covariate in the full model

Variable	Alternative effects
Individual-level covariates	
-Woman's age	Linear and/or nonlinear
-Woman's income	Linear and/or nonlinear
-Woman's age at first childbirth	Linear and/or nonlinear
-Woman's age at her first sexual intercourse	Linear and/or nonlinear
-Indigenous origin	Linear
-Education level	Linear
-Woman's age by indigenous origin	Interaction
-Woman's age by education level	Interaction
-Woman's age by age at first childbirth	Interaction
-Woman's age by age at her first sexual intercourse	Interaction
-Consent to first sexual intercourse	Linear
-Woman's age at her first sexual intercourse by condition of consent	Interaction
-Pro-gender equality attitude	Linear
Relationship-level covariates	
-Woman's age at marriage or at cohabitation	Linear and/or nonlinear
-Partner's age	Linear and/or nonlinear
-Woman's age by partner's age	Interaction
-Partner's income	Linear and/or nonlinear
-Woman's income by partner's income	Interaction
-Overcrowding	Linear and/or nonlinear
-Woman's consent to marriage or cohabitation	Linear
-Woman's age at marriage or at cohabitation by condition of consent	Interaction
-Woman's age by age at marriage or at cohabitation	Interaction
-Division of housework among household members	Linear
-Woman's level of autonomy to make decisions about her sexual life	Linear
-Woman's level of autonomy to make decisions about her professional life and use of economic resources	Linear
-Woman's level of autonomy to make decisions about her participation in social and political activities	Linear
-Social networks	Linear
-Level of social interaction reported by the woman	Linear
Community-level covariates	
-Social marginalization	Linear
-Type of community	Linear
-Women homicide rate	Linear and/or nonlinear
-Men homicide rate	Linear and/or nonlinear
-Total homicide rate	Linear and/or nonlinear
-Women's household headship	Linear and/or nonlinear
-Migration of women	Linear and/or nonlinear
-Migration of men	Linear and/or nonlinear
-Gini index	Linear and/or nonlinear
-Human development index	Linear and/or nonlinear
-Municipal functional capacities	Linear and/or nonlinear
-Women's economically active population	Linear and/or nonlinear
-Men's economically active population	Linear and/or nonlinear
-Women's political participation	Linear and/or nonlinear
-Municipality of residence	Random
-Centroid coordinates: longitude, latitude	Spatial
Societal-level covariates	
-Common crimes against women	Linear and/or nonlinear
-Common crimes against men	Linear and/or nonlinear
-Dark figure of common crimes against men	Linear and/or nonlinear
-Dark figure of common crimes against women	Linear and/or nonlinear
-Corruption	Linear and/or nonlinear
-Satisfaction with public services	Linear and/or nonlinear
-State of residence	Random

As shown in the second column of this Table 3.10, three alternative effects are considered for the 39 theoretical covariates included in the full model. First, purely linear effects are introduced for categorical covariates (See Section 2.1.1). Second, for continuous variables, as mentioned before in Section 2.1.2, instead of imposing *a priori* a particular linear form on them, we test both linear and nonlinear effects. As discussed in the literature review (Section 3.2.3), there is empirical evidence suggesting the existence of nonlinearities in some factors, such as women's age (Wilson, 2019). Finally, the introduction of interaction effects is justified for three reasons. First, the literature review indicates that some categorical variables at the individual level, such as indigenous origin and education level, alter the effect of women's age on IPV (Heidinger, 2021; Oficina de Violencia Doméstica, 2021). The same occurs with the categorical variable condition of consent with age at sexual initiation and marriage (or cohabitation). The second reason is to capture relative inequalities in age and income between the woman and her partner, as studied by Chaurasia et al. (2021), Rapp et al. (2012), and Reichel (2017). The third reason for including the interactions is the definition of the factors, which should be considered in the modeling design. This is the case of the interaction between the woman's age and age at first childbirth. Age at first childbirth depends on the value of a woman's age. Moreover, the effect of having had the first child at, say, 16 years old would be different for an 18-year-old girl than for a 50-year-old woman. Something similar happens with the interaction of a woman's age with age at marriage (or cohabitation).

Given the high dimensionality and complexity of the model in Equation 3.2, we implement the three-step methodology described in Section 2.2 consisting of the application of the boosting algorithm (see Section 2.2.1), stability selection (see Section 2.2.2), and calculation of 95% pointwise bootstrap confidence intervals (see Section 2.2.3). Empirical findings are described in the Section 3.2.6.

Implementation details

The three-step methodology for the model in Equation 3.2 is implemented as follows in this study. First, we apply the boosting algorithm to estimate the model with a step size of 0.5. Then, to determine the optimal stoping iteration we perform an empirical risk estimation with up to 10,000 iterations. The model is optimized at 3314 iterations.

Then, complementary pairs stability selection with PFER control is applied to avoid falsely selecting covariates. For this study, we use 50 complementary pairs for the error bounds and set a cutoff of 0.8. Given the number of potential predictors and their alternative effects in our model, this cutoff corresponds to a PFER with a significance level of 0.0425.

Lastly, 95% confidence intervals for the subset of effects selected as stable are calculated by drawing 1000 random samples from the empirical distribution of the data using a bootstrap approach based on pointwise quantiles.

All computations are implemented in the R package "mboost" (Hothorn et al., 2020). The corresponding code to replicate these results can be found in the Supplementary information 5.9 and is also freely available from Figshare at https://doi.org/10.6084/m9.figshare.21183271.

3.2.6 Application results

After applying the three-step methodology to the structured additive probit model in Equation 3.2, of the total of 77 alternative effects only nine of them are selected as significantly associated with emotional IPV victimization. These results are summarized in Table 3.11 and discussed in the following paragraphs according to their corresponding level of the ecological model.

Table 3.11 Selected variables associated with emotional IPV victimization

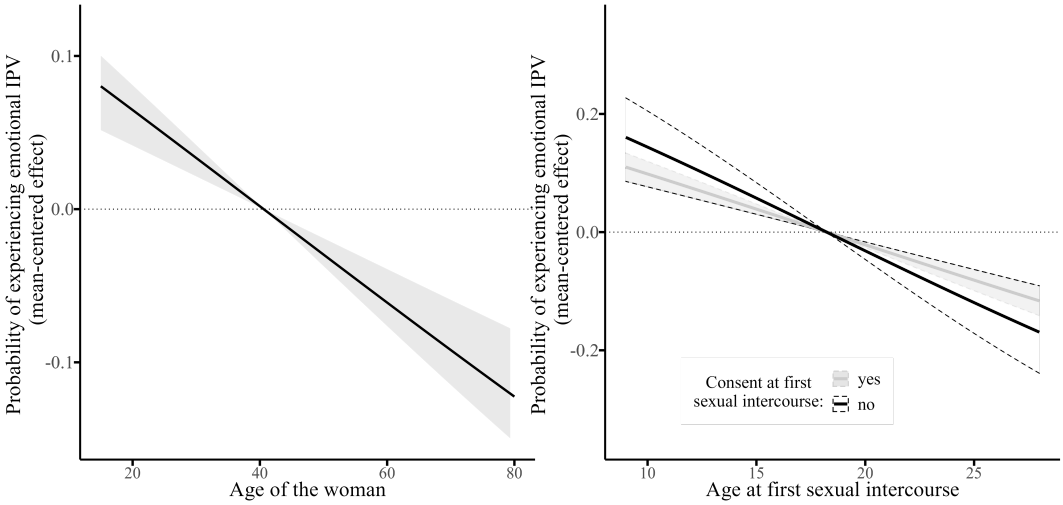
Variable	Categories	Coefficient [95% CI]
Individual-level covariates		
-Woman's age		Linear, slope: -0.003 (Fig. 3.6a)
-Woman's age at first sexual intercourse by condition of consent to first sexual intercourse	no*	Linear, slope: -0.012 (Fig. 3.6b)
	yes	Linear, slope: -0.018 (Fig. 3.6b)
Relationship-level covariates		
-Woman's age at marriage or at moving in together with current partner by condition of consent to it	no*	
	yes	Linear, slope: 0.003 (Fig. 3.7)
-Women's autonomy about her professional life and use of economic resources	low*	
	medium	- 0.1 [-0.129, -0.063]
	high	
-Social networks	low*	
	medium	0.079 [0.062, 0.097]
	high	
-Division of housework among household members	only women*	
	both	
	only men	-0.07 [-0.086, -0.054]
Community-level covariates		
-Gini index		Nonlinear, inverted U-shape (Fig. 3.8a)
-Economically active women population		Linear, slope: 0.002 (Fig. 3.8b)
Societal-level covariates		
-Prevalence of common crimes against men		Linear, slope: 0.000003 (Fig. 3.9)

Reference categories are denoted with *.

Individual-level risk factors

Two effects are found to be significant for emotional IPV at the individual level. First, regarding the effect of age on victimization, we find a linear decreasing relationship, suggesting that young women are at the most risk of victimization (Fig. 3.6a). Specifically, the risk of emotional victimization for girls around 15 years old is approximately eight percentage points higher than for women aged 40 and about 20 points higher than for women aged 80. Woman's age at first sexual intercourse is also relevant for emotional IPV (see Fig. 3.6b). Results indicate that women who had their sexual initiation at an early age are generally at higher risk of suffering emotional IPV. This effect does not differ between women who consented to their first sexual experience and those who did not.

Fig. 3.6 Effects of selected continuous covariates at the individual level



(a) Emotional IPV risk and women's age **(b)** Emotional IPV risk and women's age at first sexual intercourse by consent

Figures show smoothed mean effects with 95% empirical bootstrap confidence intervals. Coefficients express the effect on women's probability of being victims of IPV and are obtained by using the cumulative standard normal distribution centered on the mean.

Relationship-level risk factors

At the relationship level, women's age at marriage or cohabitation is positively associated with the likelihood of experiencing emotional IPV only for those who consent to it. This association is represented by a line increasing at a constant rate of 0.3 percentage points per year of age (Fig. 3.7).

Fig. 3.7 Effects of selected continuous covariates at the relationship level

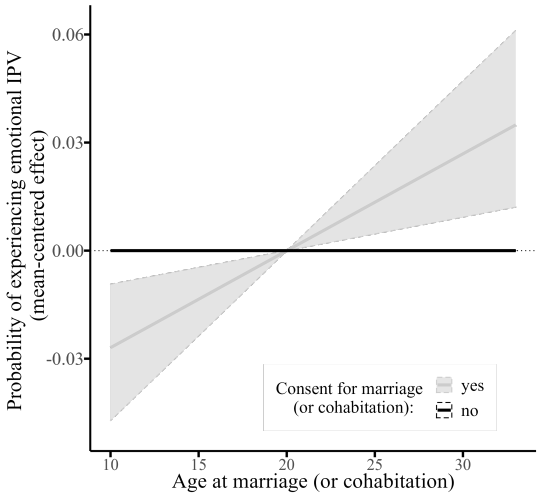


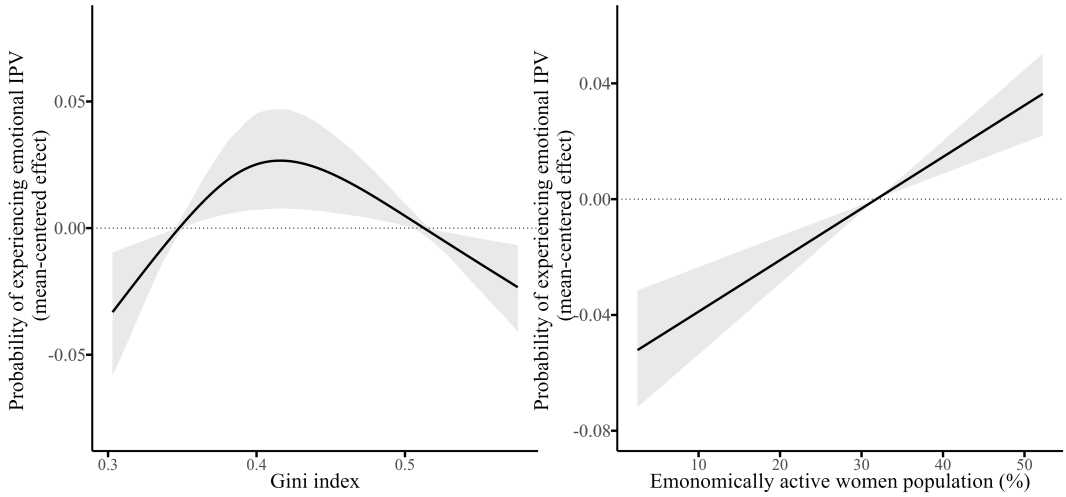
Figure shows smoothed mean effects with 95% empirical bootstrap confidence intervals. Coefficients express the effect on women's probability of being victims of IPV and are obtained by using the cumulative standard normal distribution centered on the mean.

A woman's decision-making autonomy about her professional life is also a relevant factor for emotional IPV victimization (see Table 3.11). Results indicate that compared to women with poor decision-making power, women with a medium level of autonomy are at less risk of emotional IPV victimization. No significant differences are observed between women with low and high autonomy levels. We also find that women who have a medium level of social support networks experience, on average, a higher risk of emotional IPV than those with low and high perceived social connectedness (about eight percentage points more). Furthermore, results indicate that partnered women in families in which the housework is done by the men members exhibit a risk of emotional IPV that is around seven percentage points lower than that of women in households with a different distribution of housework (see Table 3.11).

Community-level risk factors

Concerning the community level, we find that the association between economic inequality, measured by the Gini index of the municipality, and the women's likelihood of experiencing emotional IPV follows an inverted U-shaped curve (see Fig. 3.8a). Moreover, the participation of women in the community's economic activity is positively associated with IPV risks, as can be observed in Fig. 3.8b.

Fig. 3.8 Effects of selected continuous covariates at the community level



(a) Emotional IPV risk and community's Gini index

(b) Emotional IPV risk and community's economically active women population

Figures show smoothed mean effects with 95% empirical bootstrap confidence intervals. Coefficients express the effect on women's probability of being victims of IPV and are obtained by using the cumulative standard normal distribution centered on the mean.

Societal-level risk factors

At the societal level, we find that the prevalence of common crimes against men in the region is positively associated with the likelihood of women and girls experiencing emotional IPV (see Fig. 3.9). Women and girls living in regions with a prevalence rate of around 50,000 male victims per 100,000 men show a risk of emotional IPV approximately six percentage points higher than

those living in regions with a rate around the national mean of approximately 30,000 male victims per 100,000 male inhabitants.

Fig. 3.9 Effects of selected continuous covariates at the societal level

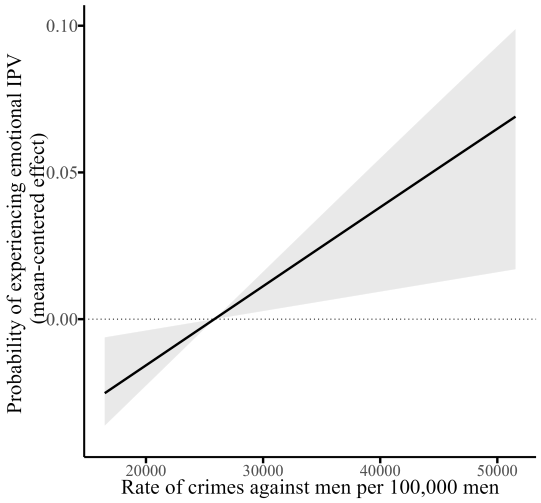


Figure shows smoothed mean effects with 95% empirical bootstrap confidence intervals. Coefficients express the effect on women's probability of being victims of IPV and are obtained by using the cumulative standard normal distribution centered on the mean.

3.2.7 Discussion of the application results

The significant factors presented in the previous section imply relationships between the selected covariates and the likelihood of emotional IPV victimization. Even though these relationships do not necessarily imply causality, they provide evidence about important aspects of emotional IPV in Mexico. In the following lines we discuss some possible explanations drawn from studies and theories presented in the literature review.

As pointed out by UNiTE Working Group (2019), Walton-Moss et al. (2005), and WHO (2012), the reasons underlying the age-victimization decreasing relationship might be related to the development of empowerment strategies and life skills throughout a woman's life.

Regarding the negative correlation between the emotional IPV risk and age at first sexual activity, this result aligns with those of other international studies (Stöckl et al., 2014) and lends support to the argument that experiences during childhood and adolescence have a major long-run impact on individuals' physical, mental, and social health. In particular, an early sexual experience is associated with many negative outcomes (Olesen et al., 2012).

Although our results regarding the positive linkage between women's age at marriage (or at cohabitation) and her probability of experiencing emotional IPV contradict previous findings (WHO, 2012), there are two potential interpretations. On the one hand, it is generally expected that women who marry at a late age have greater economic power and better social opportunities (Field et al., 2016), and this could be prompting their partners to inflict IPV in an attempt to control their resources (Bloch & Rao, 2002). On the other hand, women who marry at a late age might be "tolerating" emotional IPV to avoid being unmarried and the subsequent lingering social stigma existing in Mexico (Cuevas Hernández, 2010; Médor, 2013), and/or because of concern for their children (UNiTE Working Group, 2019) .

This result partially agrees with existing studies regarding the significant effect of women's decision-making autonomy (Mavisakalyan & Rammohan, 2021). It could indicate that when a woman has a low level of autonomy regarding her professional life and use of resources, her partner exercises dominance and control over her through emotional violence. As a woman's autonomy increases to a medium level, emotional IPV decreases because she is better placed to advocate for her rights and preferences. However, when her autonomy reaches a high level, her partner seeks to exercise his dominance and control over her and her resources via emotional IPV.

Strong social connectedness is also found to be relevant for emotional IPV. Although our results differ slightly from those of Mavisakalyan and Rammohan (2021) and Yilmaz (2018), we could nevertheless argue that for a woman at a certain level of IPV risk, as she increases her social interactions, the tensions, conflicts, and disputes with her partner initially rise, leading to a greater likelihood of victimization. After a certain level of social support networks is surpassed, the IPV risk decreases to its initial level.

With regard to the distribution of housework among the family members, we could argue that since this factor is a key gender equality indicator (Ferrant et al., 2014), in families with traditional gender roles, the housework is exclusively done by women, and this inequality is also reflected through emotional IPV. By contrast, households in which only men do the housework

seem to represent a safer place for women in terms of IPV victimization.

Our findings at the community level differ to some extent from previous results based on (UNiTE Working Group, 2019; WHO, 2012). Our results support the existence of a nonlinear relationship between the community's Gini index and IPV risks. This indicates that lower risks of emotional IPV are observed in women living in highly unequal and highly equal communities. Even though the shape of the estimated relationship differs from previous studies (Rashada & Sharaf, 2016), the results are consistent in terms of the relevance of this factor.

Results regarding the effect of the share of economically active women suggest that a greater degree of women's economic empowerment in the community's public life, in particular in job market access, could be generating tensions and conflicts in the private sphere. This may exacerbate existing gender inequalities in the context of intimate relationships, thus increasing women's IPV victimization risks.

At the societal level, the association found between the prevalence of common crimes against men and women's likelihood of experiencing emotional IPV makes logical sense. This finding is consistent with previous studies (WHO, 2012).

3.3 Examining gender inequalities in factors associated with income poverty in Mexican rural households

3.3.1 Background

Although Mexico is one of the world's 20 largest economies, 2016 official estimates indicate that around 62 million people (50.6% of the total population) still had income levels below the poverty line, of which nearly 21.4 million (17.5%) could not even afford the basic food basket (INEGI, 2016a). Furthermore, it is well known that households in rural communities are in a more difficult situation to escape from poverty since they encounter particular challenges due to a more limited access to public services (such as health and education), infrastructure, markets, job opportunities, and financial services (de Janvry & Sadoulet, 2000; Khan, 2001; Mckinley & Alarcón, 1995; Verner, 2005). Specifically for the case of Mexico, almost 60% of the total rural population (around 16.9 million people) had income levels below the poverty line in 2016, and approximately 29.2% (8.3 million) could not even afford the basic food basket (INEGI, 2016a). These figures call for the need to understand the associated factors to this higher prevalence of rural poverty to effectively diagnose the problem and to design policy interventions geared to the poorer.

Income poverty in rural Mexico is not an understudied subject. Some consistent findings can be derived from previous research on the matter regarding economic, demographic, and social factors linked to this phenomenon. Broadly speaking, there is a consensus that old-age, indigenous origin, low levels of education, overcrowded families, undernutrition, community's level of marginalization and social deprivations are associated with higher poverty levels (Hausmann et al., 2020; La Fuente, 2010; Lopez-Feldman et al., 2007; Mckinley & Alarcón, 1995; Mora-Rivera & García-Mora, 2018; World Bank, 2005). However, despite these key contributions, there are still some issues that have not been examined. On the one hand, existing studies on rural poverty mostly ignore gender, overlooking the dissimilar experiences faced by women and men in several spheres of life, such as the use of time, social networks, political and economic participation, or gender-based violence. On the other hand, most of the research is exclusively based on mean regression

models analyzing the population's average income or the expected probability of being poor, disregarding specific effects in poorer income levels.

Taking into account the abovementioned issues, in this study we aim at enhancing knowledge on rural poverty in Mexico by identifying a set of factors associated with the income-to-poverty ratio with a particular focus on detecting heterogeneous effects according to the sex of the household head. The income-to-poverty ratio, calculated by dividing the income of the household by the poverty threshold, also enables us to examine how these effects vary with the severity or depth of poverty experienced by the families. The data set used is composed of microdata with the household as a recording unit and information on 42 explanatory factors from 10 sources covering three key levels of analysis. The first level includes characteristics of the individual and its household, the second level contains the community's features, and the third one incorporates features of the region of residence (Haughton & Khandker, 2009).

Our research methodology is based on estimating a quantile regression model that allows computing specific parameters for different intervals or subsets of the distribution of the dependent variable. Particularly, two quantiles of the income-to-poverty ratio distribution are analyzed in this research, namely the corresponding poor and extremely poor households. Each poverty level is separately modeled for women- and men-headed households. In this way, we are able to examine the extent to which the effect of the explanatory variables on the income-to-poverty ratio differs by sex of the head and if these gendered effects are constant along the poverty spectrum.

3.3.2 Research questions of this study

Broadly speaking, the goal of this research is to examine, in an evidence-based way, the effect of potential risk factors on two of the lowest quantiles of income-to-poverty ratio distribution, namely the corresponding to poor and extremely poor rural families, and focusing on identifying heterogeneous effects according to the sex of the household head. Specifically, we aim at contributing to the debate on this matter by delivering key insights to answer the following relevant aspects about rural poverty in Mexico:

- Which factors at the individual, community, and regional levels are significantly associated with income poverty in rural Mexican families?

Does the effect vary with the depth of poverty? Are these factors equally affecting the income level of women and men?

- How do poverty risk factors vary throughout life? Do these effects differ across demographic groups (sex, indigenous origin, or by education level)?
- Which factors have a nonlinear effect on income poverty?
- How are the gender roles (regarding use of time, peer networks, economic and political participation, and gender-based violence) altering the risks of suffering income poverty?
- What are the economic, demographic, and political characteristics of the rural areas with greater levels of poverty? Are inequality, criminality, corruption, and quality of public services relevant to understand income poverty?

To discuss about these key issues on poverty, this section 3.2 is organized as follows. First, we give an overview of the theoretical approach used to analyze income poverty in this research. Then, we present the data and the models utilized. Posteriorly, we comment our results focusing on the identification of gender biased effects. Then, we discuss the results aiming to give some potential explanations based on previous studies and theory. Finally, we present the conclusions of this research.

3.3.3 Theory on the causes of and risk factors for poverty

The empirical evidence concerning income poverty indicates that poverty is linked to factors that can be classified into three main levels: individual and household characteristics, features of the community, and region-level characteristics (Haughton & Khandker, 2009). To present the factors found as significant across studies from different countries, we discuss some findings at each level to posteriorly examine the case of rural households in Mexico.

Individual- and household-level factors

With regard to individual and household characteristics, it has been extensively found that variables age, sex, level of education, marital status, access

to credit, and health status are associated with poverty (Haughton & Khandker, 2009). Even that these findings are not consistent across countries and regions; it can be hypothesized that young people, elders, women, individuals with poor educational achievements, married persons, and those with a bad health status tend to exhibit lower income levels. Findings supporting these hypothesis can be found in Bogale et al. (2005) and Khan (2001) for the cases of Ethiopia and a study of developing countries, respectively.

From a multidimensional perspective of poverty, it can be expected that income poverty is also linked to social deprivations such as lack of access to health services, social security, and food (Ordaz, 2009; Torres Munguía & Martínez-Zarzoso, 2020). Notwithstanding its growing importance in poverty issues, the linkage of other variables such as social networks and time use has been considerably less explored. However, overall findings show that individuals with little or no social support and those dedicating more time to domestic work are expected to be poorer (Adeyonu & Oni, 2014; Klärner & Knabe, 2019).

Community-level factors

In addition to individual and household characteristics, features at the community level are also key to study poverty. Concerning the community level, household income tends to be lower in rural settlements with poor levels of infrastructure, with lower shares of immigrant population, more vulnerable to natural disasters, marginalized, with lower shares of economic participation, and more unequal in terms of income (Baez et al., 2013; ILO, 2008). These findings have been corroborated in studies for Indonesia (Achjar & Panennungi, 2009), Latin America (de Janvry & Sadoulet, 2000), and China (Zhu & Luo, 2010).

Regional-level factors

At the regional level, the most common indicators found in previous research are related to the public sector and governance. On the whole, high levels of corruption, low quality of public goods and services, and weak governance are related to higher poverty levels (Barbier, 2012; Haughton & Khandker, 2009). Among others, such relationships have been reported by Hamilton and Svensson (2017) with data from Sudan and by Otega and Muneer'deen (2014) for Nigeria.

Previous research analyzing poverty in Mexico

Specifically analyzing the case of Mexico, prior studies have identified a set of factors whose linkage with rural poverty is widely recognized. Generally speaking, these studies have found that higher risks of rural poverty are observed in large households, in families whose head has a low educational level, indigenous, people experiencing social deprivations such as access to food or health services, and living in marginalized communities (Ordaz, 2009; Verner, 2005; World Bank, 2005). To the best of our knowledge, there are other key factors, such as social networks, time use, corruption, and gender-based violence, whose linkage with income poverty is unexplored for rural communities in Mexico, and we aim at investigating them in this research through a gender lens.

3.3.4 Data

Sources

Based on the existing studies and research, we identify the factors theoretically associated with income poverty and then we map the official sources to obtain this information for the case of Mexico.

Data on income for rural households come from the 2016 ENIGH conducted by the INEGI. The ENIGH is a nationally representative household survey conducted every two years with the aim of providing official statistical information about the income and expenditures of the Mexican families in terms of its amount, source, and distribution (INEGI, 2016a). From the ENIGH we extract data to characterize the individuals and their households.

Information on the community and regional levels is found in the 2015 Intercensal Population Survey, the 2015 CNGMD, the 2015 ENCIG, the 2016 ENDIREH, the CENAPRED, the CONEVAL, the CONAPO and the human development index developed by the UNDP. More details on these sources can be found in CENAPRED (2020), CONAPO (2016), CONEVAL (2020), INEGI (2015a, 2015b, 2015c, 2016b, 2016c), and UNDP (2019).

Dependent variable

In order to generate information for the dependent variable, income-to-poverty ratio, household income is then divided by the corresponding poverty threshold. The official criteria for defining poverty in Mexico is established by the

CONEVAL. In accordance with these criteria, a person is considered to be poor if their income is below the total cost of both the basic food basket and the non-food basket, which embraces transportation, education, health, entertainment, among others. In contrast, a person is considered extremely poor if their income is not even sufficient to cover the cost of the basic food basket (CONEVAL, 2018, 2019).

For rural communities, these two poverty thresholds were respectively set at MXN$1715.57 and MXN$ 933.20 monthly *per capita* (CONEVAL, 2020). In this way, having as reference the official poverty threshold and considering the number of family members, an income-to-poverty ratio equal to one indicates that the family is living at the poverty line. Therefore, when the ratio of income-to-poverty is less than one, the household is considered to live under poverty, and when the income-to-poverty ratio is less than 0.544 (933.20 divided by 1715.57 per household member), the household is living in extreme poverty. In other words, the income-to-poverty ratio allows us to capture how far is income of a family from the poverty line. Summary statistics of the income-to-poverty ratio are shown in the following Table 3.12.

Table 3.12 Summary statistics of the income-to-poverty ratio

	Mean	**SD**	**Median**	**Min**	**Max**
-Woman-headed households	1.46	1.12	1.17	0.00	8.19
-Man-headed households	1.54	1.41	1.13	0.00	10.90

Independent variables

The potentially associated economic, demographic, and social factors (independent variables in the model) are chosen from previous research on the matter and include characteristics at the individual, household, community, and regional levels (see Section 3.3.3).

In total, we identify 42 theoretical poverty risk factors, 15 of which correspond to the individual/household level, 20 at the community level, and seven related to the family's region of residence.

The full list of independent variables included in the full model and their summary statistics is listed in the following Tables 3.13, 3.14, 3.15, and 3.15 both for the continuous and categorical covariates according to the sex of the household head.

Table 3.13 Summary statistics of continuous covariates in the model for women-headed households

Variable	Mean	SD	Median	Min	Max
Individual- and household-level covariates					
-Head's age	53.31	16.87	52.00	21	90
-Weekly housework hours	20.94	14.41	20.00	0.00	70.00
Community-level covariates					
-Emergencies due to weather	0.85	0.73	0.67	0.00	3.00
-Gini index	0.40	0.03	0.39	0.33	0.50
-Human development index	0.72	0.06	0.72	0.56	0.83
-Municipal functional capacities	0.30	0.15	0.29	0.00	0.75
-Women-to-men ratio of housework hours	1.76	0.42	1.69	1.20	3.42
-Women's political participation	0.21	0.10	0.21	0.00	0.50
-Migration of women	0.03	0.02	0.03	0.00	0.14
-Migration of men	0.04	0.03	0.03	0.00	0.15
-Women's household headship	0.25	0.05	0.25	0.13	0.37
-Women's economically active population	0.28	0.10	0.28	0.08	0.46
-Men's economically active population	0.66	0.07	0.67	0.43	0.80
-Women working in the primary sector	0.05	0.06	0.03	0.00	0.42
-Men working in the primary sector	0.29	0.20	0.26	0.00	0.84
-Women working in the secondary sector	0.16	0.09	0.13	0.02	0.49
-Men working in the secondary sector	0.30	0.11	0.28	0.03	0.64
-Women working in the trade sector	0.22	0.04	0.22	0.09	0.32
-Men working in the trade sector	0.12	0.04	0.12	0.01	0.23
-Women working in the service sector	0.55	0.09	0.54	0.27	0.80
-Men working in the service sector	0.28	0.12	0.25	0.05	0.66
Regional-level covariates					
-Corruption	0.87	0.05	0.89	0.73	0.95
-Satisfaction with public services	0.40	0.09	0.40	0.24	0.54
-Violence against women and girls in the community	0.21	0.04	0.21	0.14	0.37
-Violence against women and girls at school	0.16	0.03	0.17	0.10	0.21
-Violence against women and girls in the workplace	0.22	0.04	0.22	0.16	0.30
-Violence against women and girls by an intimate partner	0.25	0.04	0.25	0.18	0.33
-Violence against women and girls in the family context	0.10	0.02	0.10	0.07	0.13

See Supplementary information 5.10 for definitions of independent variables.

Table 3.14 Summary statistics of categorical covariates in the model for women-headed households

Variable	Categories	N	%
Individual- and household-level covariates			
-Education level	low*	2980	67.2
	medium	1365	30.8
	high	89	2.0
-Marital status	single*	472	10.6
	open union	463	10.4
	married	816	18.4
	separated	898	20.3
	divorced	181	4.1
	widowed	1604	36.2
-Indigenous origin	yes*	1630	36.8
	no	2804	63.2
-Social networks	low*	2913	65.7
	medium	407	9.2
	high	1114	25.1
-Credit card	yes*	578	13.0
	no	3856	87.0
-Disability	yes*	837	18.9
	no	3597	81.1
-Type of household	nuclear*	2170	48.9
	single	885	20.0
	extended	1320	29.8
	other	59	1.3
-Access to food	yes*	1157	26.1
	no	3277	73.9
-Access to health services	yes*	397	9.0
	no	4037	91.0
-Dwelling with adequate quality and sufficient space	yes*	558	12.6
	no	3876	87.4
-Educational lag	yes*	2145	48.4
	no	2289	51.6
-Access to basic housing services	yes*	1846	41.6
	no	2588	58.4
-Access to social security	yes*	2556	57.6
	no	1878	42.4
Community-level covariates			
-Social marginalization	very low*	1431	32.3
	low	1219	27.5
	medium	983	22.2
	high	732	16.5
	very high	69	1.6

Reference categories are denoted with *.
See Supplementary information 5.10 for definitions of independent variables.

Table 3.15 Summary statistics of continuous covariates in the model for men-headed households

Variable	Mean	SD	Median	Min	Max
Individual- and household-level covariates					
-Head's age	48.53	15.51	47.00	21	87
-Weekly housework hours	3.83	5.59	1.00	0.00	28.00
Community-level covariates					
-Emergencies due to weather	0.83	0.69	0.67	0.00	3.00
-Gini index	0.40	0.03	0.39	0.30	0.50
-Human development index	0.72	0.06	0.72	0.55	0.87
-Municipal functional capacities	0.29	0.15	0.27	0.00	0.75
-Women-to-men ratio of housework hours	1.78	0.43	1.71	1.20	3.47
-Women's political participation	0.21	0.10	0.21	0.00	0.52
-Migration of women	0.03	0.02	0.02	0.00	0.12
-Migration of men	0.04	0.02	0.03	0.00	0.13
-Women's household headship	0.24	0.05	0.24	0.12	0.36
-Women's economically active population	0.27	0.10	0.27	0.04	0.45
-Men's economically active population	0.66	0.07	0.67	0.42	0.82
-Women working in the primary sector	0.05	0.06	0.03	0.00	0.42
-Men working in the primary sector	0.30	0.21	0.28	0.00	0.89
-Women working in the secondary sector	0.16	0.10	0.14	0.02	0.50
-Men working in the secondary sector	0.30	0.12	0.29	0.03	0.64
-Women working in the trade sector	0.22	0.05	0.21	0.09	0.32
-Men working in the trade sector	0.12	0.05	0.12	0.01	0.23
-Women working in the service sector	0.54	0.10	0.54	0.25	0.79
-Men working in the service sector	0.27	0.11	0.25	0.04	0.58
Regional-level covariates					
-Corruption	0.86	0.05	0.88	0.73	0.95
-Satisfaction with public services	0.40	0.08	0.40	0.24	0.54
-Violence against women and girls in the community	0.21	0.04	0.21	0.14	0.37
-Violence against women and girls at school	0.16	0.03	0.17	0.10	0.21
-Violence against women and girls in the workplace	0.22	0.04	0.22	0.16	0.30
-Violence against women and girls by an intimate partner	0.25	0.04	0.25	0.18	0.33
-Violence against women and girls in the family context	0.10	0.02	0.10	0.07	0.13

See Supplementary information 5.10 for definitions of independent variables.

Table 3.16 Summary statistics of categorical covariates in the model for men-headed households

Variable	Categories	N	%
Individual- and household-level covariates			
-Education level	low*	9289	62.4
	medium	5187	34.9
	high	401	2.7
-Marital status	single*	584	3.9
	open union	3391	22.8
	married	9773	65.7
	separated	470	3.2
	divorced	127	0.9
	widowed	532	3.6
-Indigenous origin	yes*	5820	39.1
	no	9057	60.9
-Social networks	low*	7852	52.8
	medium	1873	12.6
	high	5152	34.6
-Credit card	yes*	2290	15.4
	no	12587	84.6
-Disability	yes*	1995	13.4
	no	12882	86.6
-Type of household	nuclear*	10582	71.1
	one-person	1227	8.2
	extended	2963	19.9
	other	105	0.7
-Access to food	yes*	3244	21.8
	no	11633	78.2
-Access to health services	yes*	1857	12.5
	no	13020	87.5
-Dwelling with adequate quality and sufficient space	yes*	2225	15.0
	no	12652	85.0
-Educational lag	yes*	6175	41.5
	no	8702	58.5
-Access to basic housing services	yes*	6479	43.6
	no	8398	56.4
-Access to social security	yes*	9493	63.8
	no	5384	36.2
Community-level covariates			
-Social marginalization	very low*	4722	31.7
	low	3882	26.1
	medium	3092	20.8
	high	2829	19.0
	very high	352	2.4

Reference categories are denoted with *.
See Supplementary information 5.10 for definitions of independent variables.

Once collected the list of variables from the abovementioned tables, we combined them to create a unified set of information (see Supplementary information 5.11 for a detailed description of this data integration process). Posteriorly, we apply plausibility analysis, detected outliers, and removed

cases with missingness before modeling (a description of this data process can be found in Supplementary information 5.12). In sum, the final data set for the study on rural poverty is composed of information on 4,434 women-headed and 14,877 men-headed households. Each of these two sets of data contains 42 theoretical poverty risk factors describing the individual/household, community, and regional levels. These data is freely available from Figshare at https://doi.org/10.6084/m9.figshare.21183271.

3.3.5 Model specification

In this research, we utilize structured additive quantile regression models to identify how and to what extent the set of theoretical covariates is associated with the income-to-poverty ratio of the poor and extremely poor rural families in Mexico. To examine whether the linkages between the response variable and the independent variables vary according to the sex of the head, and according to the household income level, we estimate separately four additive models. Two models are applied to data on households headed by a woman and are estimated for the quantiles corresponding to the poor and extremely poor families. Analogously, the other two regression models correspond to poor and extremely poor man-headed families.

Formally expressing the abovementioned specifications, consider the dependent variable $y_{\tau i}^{sex}$, income-to-poverty ratio of observation i at quantile τ for $sex = woman, man$, according to the sex of the household head, and the set of 14 categorical $w_1, ..., w_{14}$ and 28 continuous covariates $z_1, ..., z_{28}$ (see Tables 3.13, 3.14, 3.15, and 3.16). Thus, both for women- and men-headed households, the model for the quantile τ of income-to-poverty ratio is given by:

$$y_{\tau iv} = \beta_{0\tau} + \sum_{l=1}^{14} \beta_{l\tau} w_{ivl} + \sum_{r=1}^{28} s_{r\tau}(z_{ivr}) + s_{geo\tau}(lon_{vi}, lat_{vi}) + \sum_{d=1}^{2} s_{int_d\tau}(varying_{ivd}) + \phi_{0i\tau} + \varepsilon_{\tau iv} \tag{3.3}$$

where, similar to Equation 2.3, the seven right-hand-side terms in Equation 3.3 are described below:

- $\beta_{0\tau}$ is the quantile-specific model intercept;

- $\beta_{1\tau}, ..., \beta_{14\tau}$ represent the parametric component for estimating linear effects of the 14 categorical variables (see Section 2.1.1);
- $s_{1\tau}(z_1), ..., s_{28\tau}(z_{28})$ is the model component for the 28 univariate continuous variables (see Section 2.1.2);
- $s_{geo\tau}(lon, lat)$ are the spatial effects estimated based on the geographic coordinates lon and lat corresponding to the centroid of each municipality (see Section 2.1.3);
- $s_{int_1\tau}(varying_1)$ and $s_{int_2\tau}(varying_2)$ denote the component for interaction effects (see Section 2.1.4):
  - age of the head by education level, and
  - age of the head by marital status.
- $\phi_{0i\tau}$ denotes the cluster-specific random intercept due to the hierarchical data structure, in which individual/household observations are connected to the information for the communities, and these, in turn to the regional information (see Section 2.1.5); and,
- $\varepsilon_{\tau i}$ represents the quantile-specific regression errors.

This way, the structured additive quantile regression model expressed in Equation 3.3 has a total of 75 potential effects associated to 42 theoretical covariates (see Tables 3.13, 3.14, 3.15 and 3.16) plus spatial and random effects. The following Table 3.17 lists these alternative effects included in the full model.

Table 3.17 List of alternative effects by covariate in the full model

Variable	Alternative effects
Individual-/household-level covariates	
-Head's age in years	Linear and/or nonlinear
-Education level	Linear
-Marital status	Linear
-Head's age by education level	Interaction
-Head's age by marital status	Interaction
-Indigenous origin	Linear
-Social networks	Linear
-Credit card	Linear
-Disability	Linear
-Type of household	Linear
-Access to food	Linear
-Access to health services	Linear
-Dwelling with adequate quality and sufficient space	Linear
-Educational lag	Linear
-Access to basic housing services	Linear
-Access to social security	Linear
-Weekly housework hours	Linear and/or nonlinear
Community-level covariates	
-Social marginalization	Linear
-Emergencies due to weather	Linear and/or nonlinear
-Gini index	Linear and/or nonlinear
-Human development index	Linear and/or nonlinear
-Municipal functional capacities	Linear and/or nonlinear
-Women-to-men ratio of housework hours	Linear and/or nonlinear
-Women's political participation	Linear and/or nonlinear
-Migration of women	Linear and/or nonlinear
-Migration of men	Linear and/or nonlinear
-Women's household headship	Linear and/or nonlinear
-Women's economically active population	Linear and/or nonlinear
-Men's economically active population	Linear and/or nonlinear
-Women working in the primary sector	Linear and/or nonlinear
-Men working in the primary sector	Linear and/or nonlinear
-Women working in the secondary sector	Linear and/or nonlinear
-Men working in the secondary sector	Linear and/or nonlinear
-Women working in the trade sector	Linear and/or nonlinear
-Men working in the trade sector	Linear and/or nonlinear
-Women working in the service sector	Linear and/or nonlinear
-Men working in the service sector	Linear and/or nonlinear
-Municipality of residence	Random
-Centroid coordinates: longitude, latitude	Spatial
Regional-level covariates	
-Corruption	Linear and/or nonlinear
-Satisfaction with public services	Linear and/or nonlinear
-Violence against women and girls in the community	Linear and/or nonlinear
-Violence against women and girls at school	Linear and/or nonlinear
-Violence against women and girls in the workplace	Linear and/or nonlinear
-Violence against women and girls by an intimate partner	Linear; and/or nonlinear
-Violence against women and girls in the family context	Linear and/or nonlinear
-State of residence	Random

Three alternative effects are taken into account for each of the 42 covariates considered in the full model. First, purely parametric or linear effects for the categorical independent variables (Section 2.1.1). Furthermore, for continuous covariates we proceed as described in Section 2.1.2, *i.e.* given that no functional form is imposed *a priori* to continuous variables, both linear effects and nonlinearities are considered as modeling competing options for each of them. The motivation behind this is found in existing research pointing to the existence of nonlinear effects on income of covariates such as head's age (Deyshappriya & Minuwanthi, 2020). Third, we also introduce interaction effects between head's age and categorical covariates education level and marital status. Previous studies have found that the effect of both education and marital status varies across lifetime (Torres Munguía & Martínez-Zarzoso, 2020). Moreover, it is evident that the level of education of the head depends by definition on her/his age. Similarly occurs with the interaction of age and marital status.

Model in Equation 3.3 has a large number of parameters to estimate, leading to a complex and high-dimensional setting with which tradional regression methods can not find a solution. We therefore apply the three-step methodology described in Section 2.2. Details on the implementation of this methodology applied to this study of rural poverty are described in the following Section 3.3.5. Findings are described in the Section 3.3.6.

Implementation details

We apply the three-step methodology to the model expressed in Equation 3.3 as follows. For each of the four models estimated in this research, 5000 initial boosting iterations are performed. After cross-validating to prevent overfitting the prediction accuracy of the models is optimized at the number of iterations shown in 3.18.

Table 3.18 Number of boosting iterations optimizing the models

Head's sex	**Poverty level**	**Number of iterations**
-Man	Extremely poor	813
	Poor	2846
-Woman	Extremely poor	364
	Poor	617

Once the models are fitted at their optimal number of iterations, we execute stability selection to avoid the erroneous selection of non-relevant vari-

ables. Specifically for this study on rural poverty, we use 50 complementary pairs for the error bounds and a threshold for the relative selection frequency of 0.8, which corresponds to a significance level of 0.0381.

Finally, 95% confidence intervals for the subset of effects selected as stable in step 2 are calculated by drawing 1000 random samples from the empirical distribution of the data using a bootstrap approach based on pointwise quantiles.

All computations are implemented in the R package "mboost" (Hothorn et al., 2020). The corresponding code to replicate these results is freely available from Figshare at https://doi.org/10.6084/m9.figshare.21183271.

3.3.6 Application results

From the full model expressed in Equation 3.3, a subset of 17 effects is selected as significant. Table 3.19 shows the coefficients of these factors for women- and men-headed families living either in poverty or in extreme poverty. These coefficients indicate the effect of each factor on the income-to-poverty ratio. It is important to keep in mind that this ratio measures how far or close is a family to live in poverty based on their income, and therefore coefficients can be interpreted as estimates of the effect size as a proportion of the poverty line, *i.e.* as a share of the income required to cover the cost of the household basic food basket and the non-food basket. Interpretation in the quantile context is basically the same as in other traditional approaches. For categorical covariates, parameters indicate the difference in the estimated effect of a category on the response with respect to the corresponding effect of the reference category. For example, when examining the estimated coefficients of women-headed households living in extreme poverty, results indicate that families without access to credit cards have an income-to-poverty ratio that is smaller by 0.114 units in comparison to their counterparts having credit cards. For continuous covariates with purely linear effects, the parameter indicates the change in the income-to-poverty ratio per unit change in the continuous covariate. For instance, for men-headed households living in extreme poverty, an increase of one year in the household head's age decreases the income-to-poverty ratio by 0.003 units. For continuous covariates with nonlinear effects, interpretation is best done by visualizing the corresponding figures. A comparison of the estimated coefficients from models provides a clear picture of how the covariate effect varies across the poverty spectrum and according to the head's sex.

Table 3.19 Selected variables associated with income-to-poverty ratio

Variable	Category	Extremely poor				Poor			
		Women-headed		Men-headed		Women-headed		Men-headed	
		Coef.	95% CI	Coef.	95% CI	Coef.	95% CI	Coef.	95% CI
Individual-level covariates									
-Head's age				**Linear, slope: -0.003**					
-Education level	low*								
	medium								
	high	0.594	[0.222, 1.002]	**0.319**	**[0.261, 0.415]**	0.917	[0.567, 1.825]	**0.748**	**[0.602, 0.911]**
-Marital status	single*								
	married								
	separated							**0.299**	**[0.191, 0.404]**
	divorced								
	widowed								
	open union								
-Head's age by education level	low*								
	medium	**Non-linear (Fig. 3.12)**							
	high								
-Social networks	low*								
	medium								
	high			**0.052**	**[0.036, 0.067]**			**0.093**	**[0.07, 0.116]**
-Credit card	yes*								
	no	-0.114	[-0.165,-0.055]	**-0.13**	**[-0.158, -0.109]**	-0.224	[-0.288, -0.163]	**-0.25**	**[-0.287, -0.216]**
-Type of household	nuclear*								
	single					**0.066**	**[0.027, 0.114]**	**0.667**	**[0.573, 0.770]**
	extended					**-0.11**	**[-0.149, -0.074]**		
	other								
-Access to food	no*								
	yes	**0.099**	**[0.057, 0.139]**	**0.075**	**[0.061, 0.090]**				
-Educational lag	yes*								
	no			0.033	[0.019, 0.048]	**0.094**	**[0.043, 0.144]**	0.051	[0.027, 0.075]
-Access to basic housing services	no*								
	yes	0.057	[0.023, 0.093]	**0.055**	**[0.04, 0.072]**	0.111	[0.067, 0.154]	**0.12**	**[0.097, 0.141]**
-Access to social security	no*								
	yes	0.142	[0.096, 0.187]	**0.225**	**[0.2, 0.251]**	0.159	[0.105, 0.215]	**0.31**	**[0.275, 0.344]**
-Weekly housework hours								**Non-linear (Fig. 3.13)**	
Community-level covariates									
-Gini index		**Linear, slope: -1.26 (Fig. 3.14)**		Linear, slope: -1.08 (Fig. 3.14)		**Linear, slope: -1.89 (Fig. 3.14)**		Linear, slope: -1.19 (Fig. 3.14)	
-Human development index		Linear, slope: 0.49 (Fig. 3.15)		**Linear, slope: 0.9 Fig. 3.15)**		Linear, slope: 0.75 Fig. 3.15)		**Linear, slope: 1.8 Fig. 3.15)**	
-Women's household headship		**Linear, slope: 0.65**							
-Women's economically active population		Linear, slope: 0.33 (Fig. 3.10)		**Linear, slope: 0.26 (Fig. 3.10)**		Linear, slope: 0.42 (Fig. 3.10)			
Regional-level covariates									
-Satisfaction with public services		**Linear, slope: 0.22 (Fig. 3.11)**		**Linear, slope: 0.29 (Fig. 3.11)**		Linear, slope: 0.75 (Fig. 3.11)		**Linear, slope: 0.64 (Fig. 3.11)**	
-Violence against women and girls in the community				**Linear, slope: 0.41**				**Linear, slope: 0.56**	

Reference categories are denoted with *.
Results in **bold** letters indicate that the effect varies with household income level.
Results highlighted in gray indicate that the effect varies according to the sex of the household head.
Empty cells indicate that the corresponding effect is not stable and therefore it is set to zero.

Overall, the subset of relevant effects refers to 17 variables selected as significant in at least one of the four estimated models. At the individual and household level, these are social networks, credit card ownership, type of household, access to food, educational lag, access to basic housing services, access to social security, education level, marital status, age, and weekly housework hours. At the community level, four covariates are selected, namely the Gini index as a proxy for income inequality, the human development index, women's household headship, and women's economically active population. Finally, two variables describing the region's characteristics are also found to be relevant: satisfaction with public services, and gender-based violence against women and girls in the public sphere. Linear and/or nonlinear and interaction effects are selected as functional shapes to describe these associations most appropriately. The rest of the variables are not found to be associated with income-to-poverty ratio in any of the models.

In the following lines we comment in detail on the results reported in Table 3.19. We classify the findings into two groups: risk factors whose association with the income-to-poverty ratio do not vary according to the sex of the household head (no gender bias) and those having heterogeneous effects between women- and men-headed households (gender bias). In turn, these two groups of risk factors can be divided into those whose estimated effects do not differ among poverty levels and into those risk factors with an income-poverty-level varying effect.

Income poverty risk factors without gender bias

We find no significant differences between the coefficient estimated for women-headed households and the corresponding one for men-headed households within the same poverty level in seven variables with effects selected as relevant in at least one of the four models. These variables are credit card ownership, access to food, educational lag, access to basic housing services, education level, economically active women population, and satisfaction with public services. It is important to remark, that for some variables the homogeneous effect between sexes is exclusively found in one of the two levels of poverty, but in other cases it is constantly observed for both poverty levels.

The results indicate that having a household member that holds a credit card is consistently linked to a greater income-to-poverty ratio in rural families. For extremely poor families, not having access to credit cards reduces their income-to-poverty ratio by 0.114 units in women-headed households,

and by 0.13 units for their male-headed counterparts. The magnitude of the effect significantly varies across poverty levels, but only for men-headed households (see confidence intervals in Table 3.19). The estimated parameter for the effect of not holding a credit card on the ratio of income-to-poverty of poor families is -0.224 for women-headed households and -0.25 for those headed by a man.

Concerning access to nutritious and quality food, it is relevant only for families living in extreme poverty. Broadly speaking, results suggest that extremely poor households with access to food have a greater income-to-poverty ratio than extremely poor households deprived of food. This association is 0.099 for families headed by a woman, and 0.075 for men-headed households.

Results also indicate that families whose head is lagging behind the compulsory level of education are expected to show a lower income level compared to households whose head has no educational lag. Only for households living in poverty, no evidence on gender-biased effects is found. For poor families, the estimated parameter for households with a head not lagging the compulsory education is 0.094 for women-headed families, and 0.051 for households headed by a man.

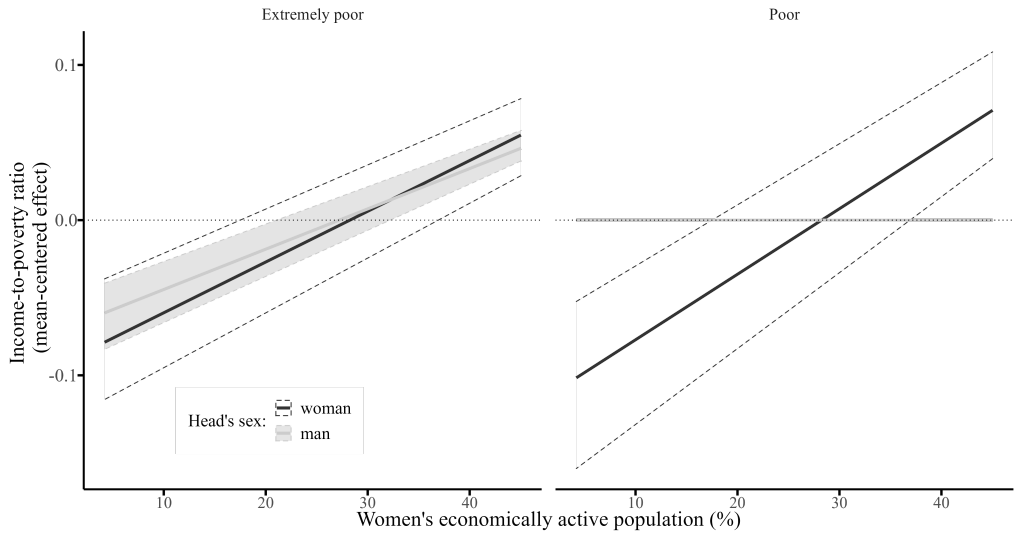
Having a house accessing to basic services is positively associated with income in rural households. The estimated parameters do not differ between women and men-headed households when keeping the poverty level constant. For the models estimated for the quantile corresponding to extreme poverty, the effects are estimated at 0.057 and 0.055 for households with a woman as head and those with a man as head, respectively. Poor households living in a house with access to basic services have an income-to-poverty ratio greater in about 0.111 for women-headed households and 0.12 for men-headed families than households deprived of basic housing services. Only for men-headed households, differences between levels of poverty are observed.

Moreover, in contrast to low and medium levels of education, having a high level of education (at least a university degree) is associated with a higher household income level in rural Mexico. For families in extreme poverty, the estimated parameter for those having a highly educated woman as the head is 0.594, and 0.319 for families with a highly educated man as the head. As one moves up the quantile distribution to the poverty level, dissimilarities are observed but only for men-headed households. The coefficients for the poor households are 0.917 for families with a woman as head and 0.748 for those headed by a man.

Concerning women's economically active population in the community of

residence, results indicate that it is positively associated with the income-to-poverty ratio for women- and men-headed households in extreme poverty and for poor households headed by a woman. No significant gender differences are found for extremely poor households. As can be seen in Fig. 3.10, for extremely poor families, the 95% confidence intervals of the estimated coefficients for the women- and men-headed households completely overlap. A one-percent rise in the share of women involved in the economic activity of the community is associated with an increase of 0.0033 in the income-to-poverty ratio of extremely poor families headed by a woman. The estimated effect of a one-percent increase in the percentage of women economically active on the income-to-poverty ratio of men-headed households in extreme poverty is 0.0026.

Fig. 3.10 Linear effects of women's economically active population on the income-to-poverty ratio by sex of the head and poverty level

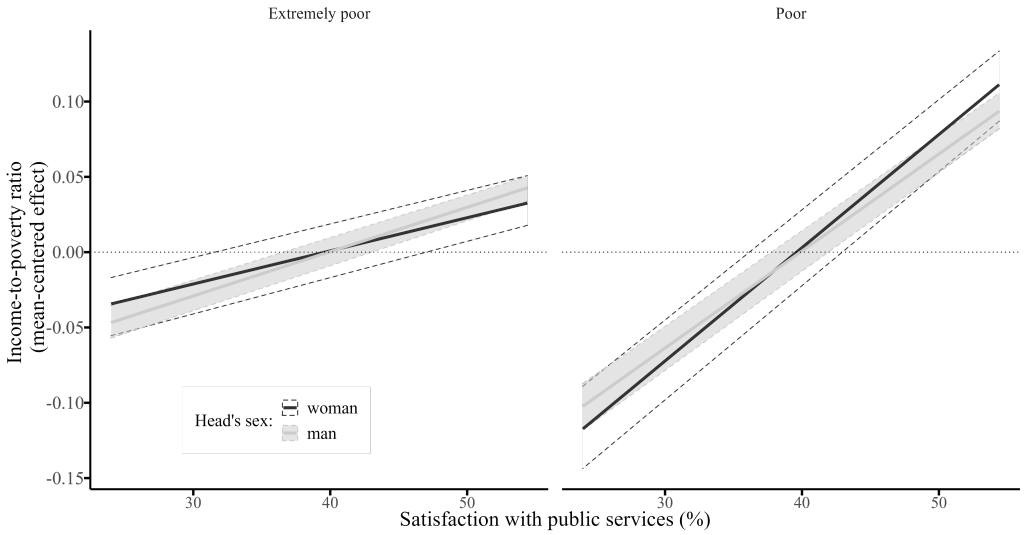


The solid lines represent the mean effects, and the dashed lines indicate 95% confidence intervals.

Finally, regarding satisfaction with public services provided in the region of residence, a positive relationship is found between this variable and the ratio of income-to-poverty of households in rural Mexico (see Fig. 3.11).

This association is selected as stable and significant in the four models. For extremely poor families, the parameter is estimated at 0.22 for households having a woman as the head and at 0.29 for households with a man as the head. For poor households, the association is 0.75 for women-headed households and 0.64 for men-headed households. As shown in Fig 3.11, both for poor and extremely poor households, the confidence intervals for the women-and men-headed completely overlap.

Fig. 3.11 Linear effects of satisfaction with public services on the income-to-poverty ratio by sex of the head and poverty level



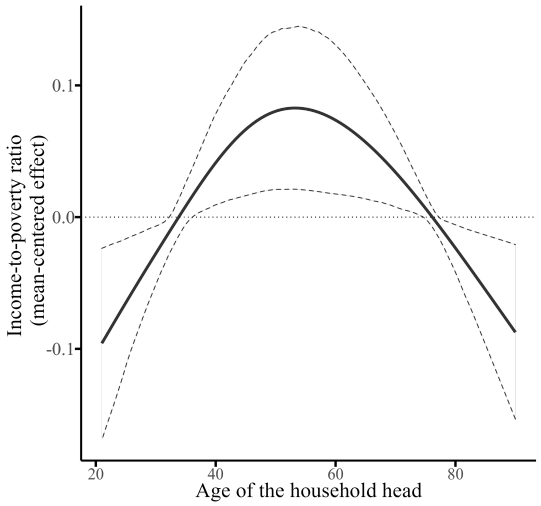
The solid lines represent the mean effects, and the dashed lines indicate 95% confidence intervals.

Income poverty risk factors with gender bias

As observed in Table 3.19, a total of 13 covariates are found to have significant gendered effects in at least one of the models estimated, *i.e.* keeping poverty level unchanged, confidence intervals of the parameters estimated for the women- and men-headed households do not intersect. For the variables educational lag, age, and the interaction of age with medium level of education, unequal gender effects are observed only for extremely poor families.

At this poverty level, the association between education lag and the ratio of income-to-poverty is only significant for men-headed families. This parameter is 0.033 for households with a head not lagging behind the compulsory education level. Age has a significant effect only for men-headed households in extreme poverty. In particular, as the age of the head increases by one year, the household income-to-poverty ratio of these rural families sinks by approximately 0.003 units. As shown in Table 3.17, interaction effects of education level and age of the head are considered as a modeling alternative. In this regard, only the age varying effect of families with a woman as head having a medium level of education is selected as relevant, and an inverted U-shaped curve describes its correlation. This means that these rural households experience lower income levels in the youngest and oldest ends of the age spectrum, reaching a maximum between approximately 50 and 60 years (see Fig. 3.12).

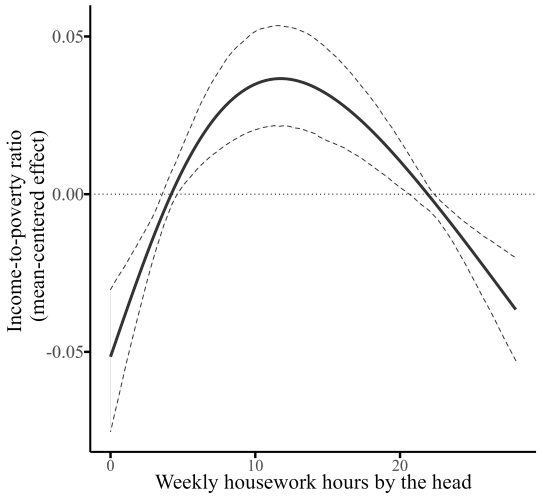
Fig. 3.12 Age-varying effects of education on the income-to-poverty ratio for extremely poor rural households headed by a woman with a medium level of education



The solid lines represent the mean effects, and the dashed lines indicate 95% confidence intervals. A medium level of education specifies that the head has a minimum of secondary education and a maximum of high school level education.

Variables with gender effects found to be significant only for the poor are the type of household, marital status, weekly housework hours, and women's economically active population in the community. In particular, in rural communities in Mexico, greater income-to-poverty ratios are expected in single households. For women-headed households in poverty, the expected income-to-poverty ratio of single households is greater in approximately 0.18 units (distance between their parameters, 0.066 and -0.11) in comparison to extended households (those composed of a nuclear family group and other family members, such as aunts, uncles, grandparents, cousins, etc.). Moreover, it is greater in 0.066 units in comparison to nuclear families and other household structures. For poor households headed by a man, the difference between the income-to-poverty ratio for single families and the rest of household types is approximately 0.667, which is almost ten times the corresponding parameter estimated for their women-headed counterparts. Regarding marital status, it is selected as influential only for men-headed households living in poverty. Specifically, families with a separated man as head show a greater income-to-poverty ratio. The coefficient of this linkage is 0.299. At the individual / household level, the linkage between income-to-poverty ratio and weekly hours doing housework is selected as relevant only in the model for men-headed households living at the poverty line (see Fig. 3.13). For these families, the linkage is represented by an inverted U-shaped curve indicating that households whose head spends less than 5 hours a week or more than 20 are associated with lower income levels. About the effect of women's economically active population in the community of residence, results suggest a gender effect indicating a positive association with the income-to-poverty ratio for women-headed poor households.

Fig. 3.13 Effects of weekly housework hours by the head on the income-to-poverty ratio for poor rural households headed by a man



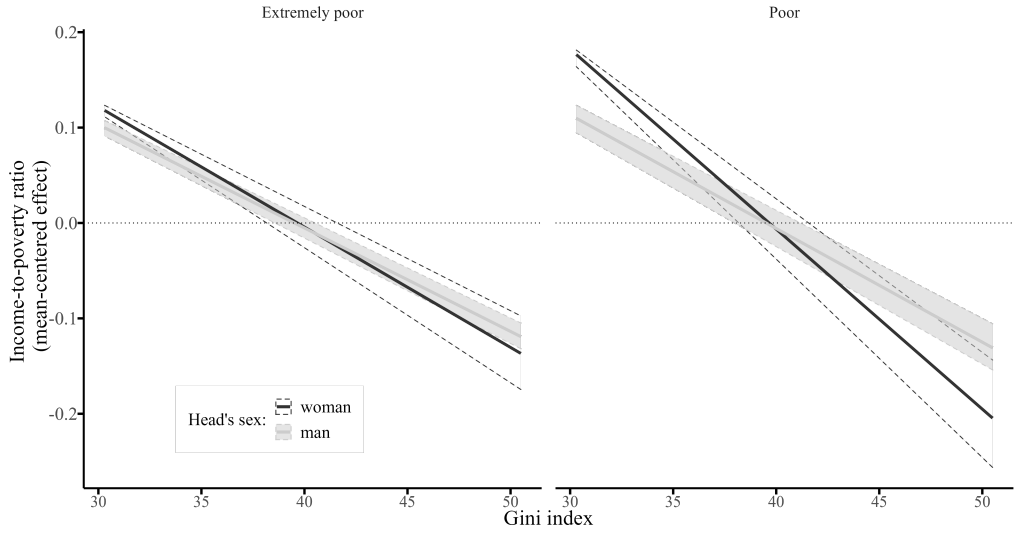
The solid lines represent the mean effects, and the dashed lines indicate 95% confidence intervals.

In addition, we also identify a subset of factors with an uneven effect on income according to the head's sex observed both in the poor and extremely poor households. Variables having this gendered effect are social networks, access to social security, Gini index, human development, and gender-based violence against women in the public sphere. About the perception of social networks, contrary to the observed in families with a woman as head, it is found that for men-headed households having a high degree of connectedness with other people is linked to a greater income-to-poverty ratio compared to families whose head has a medium or low degree of social networks. It is key to highlight how this effect varies with income level. For extremely poor families, the estimated coefficient is 0.052, whereas the effect is significantly greater for poor households, 0.093.

Findings also indicate that access to social security is linked to a greater income-to-poverty ratio. For extremely poor families, the parameters are 0.142 for women-headed households and 0.225 for men-headed families. This effect is greater for households living in poverty, whose estimated parame-

ters are 0.159 for households headed by a woman and 0.31 (3.16 standard deviations) for households headed by a man. The Gini index of the community of residence is selected as a relevant variable in all four models (see Fig. 3.14). In general, as income inequality in the municipality decreases, the household income-to-poverty ratio goes up. For extremely poor households, this association is estimated at -1.26 and -1.08, respectively, for women- and men-headed households. For poor households, the corresponding effect for families headed by a woman is -1.89, and for those headed by a man, the coefficient is -1.19. Even if all the parameters point to a negative association with income poverty, gender differences are observed in communities with the lowest levels of income inequality, in which the effect for women-headed households is expected to be larger in comparison to households headed by a man.

Fig. 3.14 Linear effects of Gini index on the income-to-poverty ratio by sex of the head and poverty level

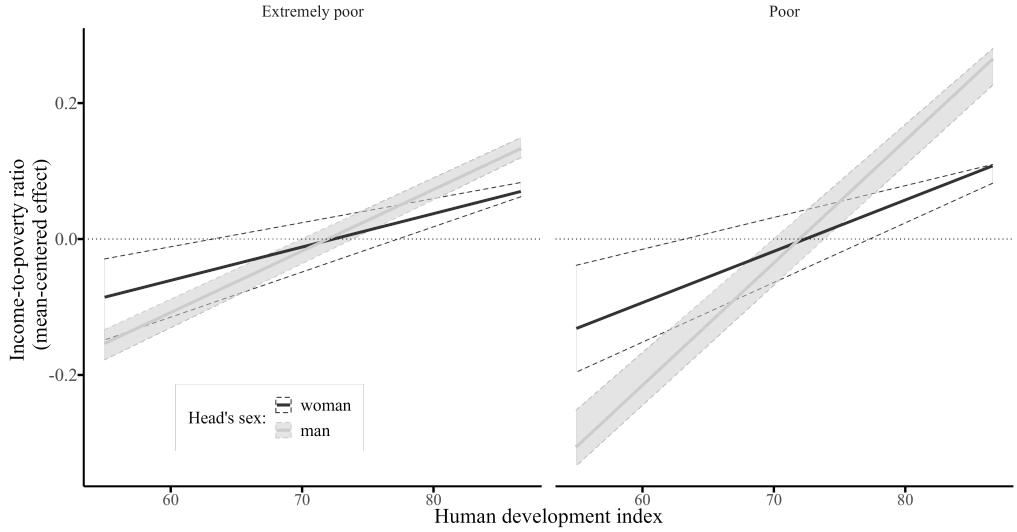


The solid lines represent the mean effects, and the dashed lines indicate 95% confidence intervals.

Moreover, the human development index of the community is also found to be stable and significant for all four groups considered. However, contrary

to the linkage found between the Gini index and the income-to-poverty ratio, the estimated parameter for the association of the human development index in the four models is linearly increasing (see Fig. 3.15). Specifically, for families living in extreme poverty, the coefficient for those headed by a woman is 0.49, and for the households headed by a man the coefficient is 0.9. For households living in poverty, in those whose head is a woman the coefficient is 0.75, and for the families with a man as head the association is estimated at 1.8 (see Table 3.19). Gender differences against women-headed families are observed in municipalities with the highest indexes of human development. In particular, there are also gender inequalities in the communities with the lowest levels of human development for poor households.

Fig. 3.15 Linear effects of human development index on the income-to-poverty ratio by sex of the head and poverty level



The solid lines represent the mean effects, and the dashed lines indicate 95% confidence intervals.

Women's household headship at the community level only shows a significant effect in the model for women-headed rural households living in extreme poverty. In particular, an increase of one percent in the share of people living in women-headed households is associated with an improvement of 0.0065 in

the family's income-to-poverty ratio (see Table 3.19). At the regional level, the variable gender-based violence against women and girls in the state of residence is stable and significant only for the income-to-poverty ratio of men-headed households. Specifically, an increase of one percent in the percentage of the women's population who was a victim of gender-based violence in the public sphere (perpetrated by a friend, an acquaintance or a stranger with whom the victim has no family nor intimate relationship, the perpetrator is not her co-worker nor her schoolmate) in the last 12 months is associated with a rise of 0.0041 in the household income-to-poverty ratio for men-headed families. This effect is greater for households headed by a man living in poverty, and it is estimated at about 0.0056 for a growth of one percent in the covariate.

3.3.7 Discussion of the application results

All the risk factors presented in the previous section imply statistical relationships between the selected variables and the household income-to-poverty ratio, and even though they do not necessarily imply causality, they provide evidence about key aspects for the studies on poverty in rural Mexican households, and some potential explanations can be derived based on existing studies and theories.

As previous research on poverty has found, social networks can help people get a job, provide financial assistance, help in childcare, influence economic decisions such as investment or expenditures, and impact on income (Skoufias et al., 2010). It is important to note that a lack of social connectedness may be an effect of poverty and a consequence. Just as the social, emotional, and financial support from close friends works as a strategy for coping with poverty, lower income levels also reduce the possibility of socializing. It is of special interest the existence of gender inequalities in the effect of social networks on income. The fact that no significant linkage is found for women-headed households could be attributable to gender differences in the structure and composition of the social networks. In Kim (2014), the authors found that women's social networks consist of mostly relatives and female neighbors, while the social networks of men are mainly formed by job- and business-related acquaintances. This could also explain our results in the case of rural Mexico.

Regarding the significant association of holding a credit card with the income-to-poverty ratio, this result is in accordance with previous studies

for other countries (de Janvry & Sadoulet, 2000; Haughton & Khandker, 2009). There are two potential interpretations of this fact. On the one hand, there might be a causal impact of having access to financial services on income through the application of economic resources from credits directed at entrepreneurial investment, smooth consumption, protection against income or price shocks, and resource allocation. On the other hand, having an insufficient income is at the same time a constraint for accessing financial services. About the type of household, one-person households show the greatest income-to-poverty ratio compared to nuclear, extended, and other family structures. Our results could suggest that the larger the number of dependents per family, the lower the income, which matches well with Verner (2005).

There is also a clear indication that the multiple dimensions of poverty are interlinked in rural communities. The significant correlation between access to food and income poverty may reveal a two-way relationship. First, consuming a nutritious diet requires allocating enough money for buying adequate food in quality and quantity, which can be a challenge for households struggling with income-related adversities. Second, consumption of nutritious food helps to maintain good health, and in turn, it improves the ability of individuals to join the labor force, obtain a better job or achieve higher productivity.

Not holding the mandatory educational level (education lag) is another aspect of multidimensional poverty associated with income for rural poor and extremely poor families. Although more detailed statistical information is required to inquire into the reasons for this result, as stated in *Priorities and strategies for education: A World Bank review* (1995), a potential explanation could be that completing compulsory education increases income because this education level is associated with a fertility rate reduction and a health status improvement through an increase in contraceptive use and a delay of the age of marriage (or age of first pregnancy), which in turn is linked to a higher productivity.

Having a house with access to basic services is positively associated with income in rural households, and this is equally affecting both women- and men-headed households. These results share similarities with the findings from Haughton and Khandker (2009). Housing represents a major charge on income. Given that the poor and extremely poor families have an income that is not enough to get an affordable and decent house, the poorest of the poor could be being forced to live in places with more marginalized circumstances

such as a lack of access to services. Simultaneously, these conditions could be affecting their health, education, access to work, and productivity.

Our results also support the existence of a positive relationship between income and access to the social protection system. This association could indicate that to have access to social security services, the head of the family must have a formal working contract, and formal jobs in Mexico are usually better paid than non-formal jobs (INEGI, 2021). It is worth noting that the results also point to gender inequalities against women-headed households, which could be corroborating the fact that in Mexico women face greater difficulties in entering the formal job market than men (INEGI, 2019).

A commonsense result is achieved when analyzing the association of education level with the income-to-poverty ratio. Consistent results highlight that families whose head completed at least university have a higher income. Our results also indicate that for families having a woman head with a medium level of education, education has an age-varying effect described by an inverted U-shaped curve. For men-headed families, the relationship is described by a downward-sloping line. The form of these shapes reflects how the income evolves heterogeneously over the life cycle, maybe due to differences in working productivity.

As expected, our results confirm significant differences according to the marital status of the head. However, unlike other research in this area indicating that lower-income levels are observed in households that experienced a dissolution (Haughton & Khandker, 2009; McManus & DiPrete, 2001), we find a greater income-to-poverty ratio for households headed by a separated man. Remarkably, this effect is not shown for families with a divorced head. A possible explanation for this may be that when separating from their partners, men have an increase in their household income as a result of a decrease in the number of dependents and because in contrast to divorced heads, there is no legal judgment dissolving the union between the male head and his ex-partner, and therefore the legal responsibilities for them, such as financial support for dependents, are not delineated.

Another interesting result on gender issues in this research is weekly time spent by the head on housework. This covariate is found to be significant for poor households headed by a man. Our results substantiate previous findings for other countries (Adeyonu & Oni, 2014; Killewald & Gough, 2010). It is logical to analyze this association as a trade-off between the time devoted by the head to paid and unpaid work and as a trade-off of time among the family members. Our results show that when all the work is done by the rest of the

household members, the income-to-poverty ratio is low, maybe because the head is the only member with time available to engage in a paid activity. As the head expends more time on housework, income raises possibly because more members spend less time on housework but more time on the market work. However, after a maximum point (between 10 and 15 hours per week), the head cannot increase his housework time without decreasing the income-to-poverty ratio of the family.

Regarding the characteristics of the community of residence, findings indicate that households in more unequal rural communities tend to exhibit lower income levels. Similarly, greater income-to-poverty ratios are observed in municipalities with the best levels of human development. These correlations could be revealing that the favorable living conditions of the communities (income equality and human development) have a positive effect on the income of the poor and extremely poor households (Lakner et al., 2020). At the same time, these associations can also reflect the residential decisions of the households: families with enough money decide to move to municipalities with better living conditions, while those with the worst income levels remain in the communities with the worst living conditions.

Regarding the correlation of women's household headship in the municipality of residence with the income-to-poverty ratio, we find a particularly intriguing result. It is well known that in patriarchal societies, assigning a woman as a head is an unusual situation frequently linked to lone-parent households or childbearing outside the marriage, so it can be expected that in municipalities with these adverse circumstances against women, families headed by a woman have on average a lower income (CONAPO, 2016). In contrast to this general assumption, we find that after controlling for the individual-, community- and regional-level characteristics, women's headship in the community and income-to-poverty ratio have a positive relationship in extremely poor households headed by women. Apart from residential decisions of women-headed families to live in municipalities where more women are acknowledged as heads of the household, it could also indicate that in these communities, women have greater well-being and empowerment that impacts on income. Alternatively, it can also suggest that they are receiving remittances from the partner.

Moreover, our results indicate that as the share of women involved in economic activity of the rural community rises, higher income levels are observed not only for women-headed households but also for men-headed households in extreme poverty. This could suggest that the inclusion of women in the eco-

nomically active population helps address labor market imbalances in rural communities, expands the working-age population, or contributes to boosting the human capital, which impacts household income. It could also be that these families decide to reside in communities where the women have greater employment opportunities.

The examination carried out in this research also reveals that the quality of the public provision of goods and services in the region of residence is positively associated with the income-to-poverty ratio. On the one hand, it can indicate that families being the poorest of the poor tend to have a residence in regions with lower quality of public services (which are likely to have a lower cost of living), while those who have an income enough to afford it reside in a region with better provision of public services. On the other hand, it may suggest that the provision of public services impacts the income of the households via an improvement in their quality of life (Hamilton & Svensson, 2017).

Finally, another interesting result on gender issues at the regional level is the linkage between gender-based violence against women and the income-to-poverty ratio. As described in the previous section, higher income levels are associated in rural areas with increases in the share of women victims of gender-based violence in the public sphere. It is particularly important to observe that this correlation is uniquely relevant for men-headed families. Given that it is well known that gender-based violence is the result of the exercise of men's power over women and girls, a probable explanation is that male heads living in a family with an income that can afford to move to another state, seek to live in communities where they can exercise their domination. A more detailed analysis of this matter is outside the scope of this research and therefore left for further research.

4. Conclusions

"It is not knowledge, but the act of learning, not possession but the act of getting there, which grants the greatest enjoyment."
Carl Friedrich Gauss

In the present research we used structured additive regression models to conduct a comprehensive analysis on three relevant topics in the field of development economics, namely use ot time, violence against women and girls in the context of intimate relationships, and income poverty. To that end, we created data sets from official information on Mexican households, incorporated a large number of theoretical factors, and proposed different modelling alternatives for the covariates, including linear, nonlinear, and interaction effects.

Given the complexity and high dimensionality of the resulting data settings, traditional approaches could not compute parameter estimation and then we applied a three-step methodology. As a first step, we implemented the component-wise gradient boosting algorithm, which has the advantage of combining estimation with automatic variable selection and model choice processes. This is fundamental in the context of structured additive models since it allowed us to leave *a priori* the functional shape of the relationship between the response and the continuous covariates unspecified, to intro-

duced a large number of potential alternative effects, and then decide for one in a fully data-driven *modus operandi*. Subsequently, we utilized stability selection in order to avoid the erroneous selection of non-relevant covariate effects. Since boosting yields to point estimates, we aplied a subsampling strategy based on bootstrap to obtain standard errors and calculate confidence intervals.

From a statistical standpoint, this strategy helped us to overcome eight common issues found in regression models in development economics:

- To deal with different types of response variables (continuous, categorical, etc.).
- The inclusion of potential nonlinear (or even *a priori* unknown) effects of continuous covariates on the response.
- To deal with a hierarchical data structure, in which individual observations are connected to the information for the communities, and these, in turn to the regional information.
- To account for spatially correlated observations.
- To explore interaction effects between a categorical and a continuous covariate.
- To examine interaction effects between two continuous covariates.
- To perform estimation with automatic identification of significant covariates (variable selection) and determination of the functional form of their linkage with the dependent variable (model choice).
- To avoid multicollinearity problems.

Probably, the greatest limitation of this three-step methodology is that it becomes computationally burdensome and expensive, and therefore time-consuming. By way of example, the estimation of the model for women-headed households living in extreme poverty in Section 3.3 lasted for about three days using the Gesellschaft für wissenschaftliche Datenverarbeitung mbH Göttingen (GWDG) RStudio server.

From an empirical perspective, the method applied led to new and original meaningful insights regarding our applications. In the research on time to

unpaid domestic work (see Section 3.1) we dealed with a continuous response variable and a cross-sectional data structure. In this study we aimed to identify the factors explaining the gap in time allocation to unpaid housework among partnered women and men in urban Mexico in 2020. Our theoretical framework considers gender inequalities as the result of the interrelationship and convergence of multiple factors describing the women, their situation within the relationship, and their families. To appropriately utilize this theoretical approach and account for the complexity of studying gender issues, such as time use among couples, we created a data set with more than 16,100 observations and 30 theoretical covariates.

All in all, a subset of seven variables was found to have a stable and significant effect on the gap of time to housework between women and men. Regarding the individual characteristics of the women these variables are: age, weekly paid working hours, and education level. About the relationship features, the selected variables describe the woman's relationship with her partner, namely woman's weekly leisure hours by partner's weekly leisure hours and the contribution of the woman's labor income in total couple's labor income. On the characteristics of their families, the gap of time to housework is found to be associated with the number of children in their household.

Our findings not only yield evidence about factors that were either hitherto unknown for the case of Mexico, understudied, or that have not yet been tested empirically, but they are also significant for the design of public policies. Accordingly, five key contributions were achieved by this research. First, we corroborated the existence of an uneven distribution of time to housework against women, in particular, women spent about 20 hours per week more than their men partners in unpaid domestic work (see Table 3.1). Second, there is an unequal intra-household decision making power to the detriment of women's time use observed in the trade-off of the time allocation among activities (paid work, unpaid housework, and leisure) and between the women and her partner. Being time a limited resource, the woman cannot increase her time spent on paid work or leisure without decreasing her time to housework, and the partner increases his time to leisure at the expense of an increase of woman's time to housework. Third, women's economic empowerment, captured by income from labor and education level, is relevant for achieving a greater equality between women and men in terms of a better time distribution. Fourth, we identified a subgroup of the women's population facing particular disadvantages against their time devoted to unpaid

housework, namely those with a low level of education, having no income from labor, and living in a family with children. Finally, women's situation relative to her partner is also associated to time allocation to unpaid housework. This way, there is an influence of the time spent on leisure by each of them, and their income from labor.

In our second study, developed in Section 3.2, we move to analyze a binomial response variable and a hierarchical data structure. Here, we aimed to identify the risk factors for emotional IPV against women and girls with children in Mexico. Our theoretical framework is the ecological model, which considers IPV as the result ofthe interaction and convergence of multiple social, demographic, economic, political, and cultural factors at four interrelated levels: individual, relationship, community, and society. To properly apply the ecological approach and account for the complexity of IPV in our analysis, we integrated a data set containing 35,004 observations and 39 theoretical covariates plus spatial and random effects with information from ten official sources. Information from the ENDIREH allowed us to characterize the individual and relationship levels. Data from the other 10 sources (including surveys, censuses, and administrative records) was used to describe the community and society in which the IPV occurs.

The main results confirmed the importance of incorporating factors at the four levels of the ecological model, rather than restricting the analysis to only the individual and relationship levels, as done in most previous research. Moreover, we found evidence of linear, nonlinear, and interaction effects describing the links between the analyzed factors and emotional IPV.

At the individual level, we found that young women and/or those who had their first sexual intercourse during childhood face a higher risk of suffering from emotional IPV. At the relationship level, women who married (or moved in together with a partner) late in life, who had a low or a high level of autonomy, who perceived a medium level of support from social networks, and/or who lived in a household in which women do all or part of the housework have a higher likelihood of emotional IPV victimization. Protective factors related to community characteristics are high-income inequality or high-income equality, and/or a low level of women's economic participation. A high prevalence of common crimes against men is associated with higher IPV victimization risks at the societal level.

These findings are also significant for public policies. In this respect, four key contributions were made by this paper. First, by examining the factors at the individual and relationship levels, we were able to identify some spe-

cific risk subgroups of the women population that are generally overlooked; namely, those who had their first sexual intercourse during childhood and women who got married (or moved in together with a partner) late in life. Strategies against IPV should focus on these at-risk groups.

Second, the results about women's autonomy and social support networks indicate that interventions aiming to promote women's social and economic empowerment should be accompanied by specific measures to protect women from violence.

Third, even if public policies already seek to promote income equality and women's economic participation in the community, our findings suggest that these policies should incorporate a gender component regarding IPV, with a particular focus on communities that have a Gini index of around 0.4 and in which a large share of women are economically active.

Finally, anti-crime policies in regions with a high incidence should include programs that also seek to reduce the risk of emotional abuse occurring in the context of intimate relationships.

In our third study developed in this thesis in Section 3.3, we applied our research strategy to study the distributions of a continuous response variable in a hierarchical data setting. This research aims to identify a set of relevant variables associated with the income-to-poverty ratio in rural Mexican households. We emphasize finding the extent to which the effect of the significant factors differs between women- and men-headed households and how these gendered effects vary according to the depth of poverty experienced by the families.

To achieve this goal, we constructed a cross-sectional data set containing information on 4,434 women-headed and 14,877 men-headed households to which we incorporate 42 variables at the individual/household, community, and regional levels, from 10 different data sources. This data set is used to estimate four additive quantile models, which allowed us to compute specific parameters for different quantiles of the income-to-poverty ratio distribution. In particular, two models were applied to data on households headed by a woman and were estimated for the quantiles corresponding to the poor and extremely poor families. Similarly, the other two models corresponded to poor and extremely poor man-headed families.

Based on the association of the considered variables with the income-to-poverty ratio, the results presented herein allowed us to distinguish two different main types of effects. First, we identified a subset of variables whose significance is consistent for poor and extremely poor families, but their

effect on income was not statistically different between women- and men-headed households. These variables were credit card ownership, access to basic housing services, education level, and satisfaction with public services.

Second, our results also identified significant differences between women- and men-headed families concerning the effects of several variables on the income-to-poverty ratio for poor and extremely poor households. Variables belonging to this group were social networks, access to social security, Gini index, human development, and gender-based violence against women in the public sphere. More importantly, for social networks, access to social security, and gender-based violence against women in the public sphere, the uneven effect between the sexes grows as family income goes from extreme poverty up to the poverty level.

Broadly speaking, these results have key implications on the study of income poverty in rural Mexico through a gendered lens. By controlling by a large set of factors at the individual/household, community and region levels, our results helped us to underscore the circumstances in which women- and men-headed households face particular disadvantages. In this regard, we detected some households, traditionally overlooked, that may experience even worse poverty levels. These are, among others, households headed by an older man, families having a younger or older woman head with a medium level of education, men-headed households lacking social networks, and extended households headed by a woman. Differently, is it worth noticing that having a highly educated woman as head of the household seems to be related with lower severity of poverty. This result emphasizes the importance of women's education as a mean of fighting poverty in rural areas.

Even that all the selected relevant covariates presented in the previous sections provide evidence about important aspects for the studies on time use, IPV, and income poverty, they exclusively imply statistical relationships between the independent variables and the respective response variables. In the following, we plan to use the three-step methodology to analyze causal effects on these topics. Moreover, further research should also consider the study on time use to leisure and housework, other forms of IPV such as sexual, physical, and economic, and to consider other poverty indicators, such as those related to multidimensional poverty, or even analyze other distributional parameters of the response, including the scale and shape.

5. Supplementary information

5.1 Implementation details for Introduction

5.1.1 Code for replicating Fig.1.1

```
### Packages ###
if(!require("ggplot2")) install.packages("ggplot2")
if(!require("ggthemes")) install.packages("ggthemes")

### Data
load("DataFig1.RData")

###
mytheme <- function(base_size = 26, base_family = "serif") {
 theme_bw(base_size = base_size,
      base_family = base_family) %+replace%
  theme(
   strip.background = element_blank(),
   #   strip.placement = "outside",
   strip.text.x = element_text(size = 17),
   strip.text.y = element_text(size = 20, angle = 270),
   legend.title.align = 0.5,
   legend.position = "right",
   legend.direction = "vertical",
   legend.title = element_text(size = 22),
   legend.text = element_text(size = 20),
   legend.key = element_blank(),
   axis.text.x = element_text(size = 17),
   axis.text.y = element_text(size = 19, hjust = 1),
   axis.ticks = element_line(colour = "black"),
   axis.title.x= element_text(size = 22),
   axis.title.y= element_text(size = 22, angle = 90),
   panel.background = element_blank(),
   panel.border = element_blank(),
   panel.grid.major = element_blank(),
   panel.grid.minor = element_blank(),
```

```
   panel.margin = unit(1.0, "lines"),
   plot.background = element_blank(),
   aspect.ratio = 1,
   plot.margin = unit(c(0.5, 0.5, 0.5, 0.5), "lines"),
   axis.line.x = element_line(color = "black", size = 1),
   axis.line.y = element_line(color = "black", size = 1)
  )
}

addline_format <- function(x,...){
 gsub('\\s','\n',x)
}

### scatterplot
Fig1 <- ggplot(saltillof[, c("ictpc", "edad_jefe")],
        aes(y = ictpc, x = edad_jefe)) +
 geom_point(size = 2.5) +
 xlab("Age␣in␣years") +
 ylab("Total␣household␣per␣capita␣income␣\n␣(in␣thousands)") +
 mytheme()

ggsave(Fig1, filename = "Fig1.png", dpi = 300,
    width = 15, height = 15, units = "cm")
```

5.2 Metadata for the data used in Section: Understanding gendered inequalities in time allocation to unpaid housework among partnered women and men in Mexico

The following are the metadata describing the attributes needed to use and understand the data utilized in the analysis developed in Section 3.1. The full data set can be found in the file called "database_timeuse.RData" and is freely available from Figshare at https://doi.org/10.6084/m9.figshare.21183271.

Variable	Description
Individual characteristics of the women	
-Woman's weekly paid working hours:	Time in hours spent on paid work by the woman per week. **Type:** continuous. **Name in the database:** *hor_1* **Source:** INEGI (2020)
-Woman's weekly leisure hours:	Time in hours spent on leisure by the woman per week. **Type:** continuous. **Name in the database:** *hor_8* **Source:** INEGI (2020)
-Woman's age:	Age in years of the woman. **Type:** continuous. **Name in the database:** *edad* **Source:** INEGI (2020)
-Woman's income-to-poverty ratio:	The income-to-poverty ratio is calculated by dividing the labor income of the woman by the poverty threshold. The official criteria for defining poverty indicates that a person is considered to be poor if their income is below the total cost of both the basic food basket and the non-food basket, which embraces transportation, education, health, entertainment, among others. For urban communities, the poverty threshold was set at MXN$ 3559.88 monthly *per capita*. **Type:** continuous. **Name in the database:** *ipovlab* **Source:** INEGI (2020)
-Educational lag:	Reported status of the educational lag of the woman. This variable indicates if the woman is lagging behind the compulsory level of education according to her age. **Type:** categorical. *"yes"* a woman has an educational lag if she was born before 1982 and has not yet completed the elementary school level; or, if 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 (2021)

Variable	Description
-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 (2021)
-Education level:	Degree of formal education level completed by the woman. **Type:** categorical. *"low"* if the maximum completed level by the woman is primary education; *"medium"* if the woman has minimum secondary education and a maximum of high school; and, *"high"* if the woman has completed at least a university degree. **Name in the database:** *nivelaprob* **Source:** INEGI (2020)
-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 (2020)
-Social networks for care:	Degree of perception of the woman on the easiness to obtain support for care from social networks in three hypothetical circumstances: care due to illness, to be accompanied to a medical appointment, and child care assistance. **Type:** categorical. *"low"* if obtaining support from social networks in the majority of hypothetical situations is perceived by the woman as difficult or impossible; *"high"* if obtaining support from social networks in the majority of hypothetical situations is perceived by the woman as easy or very easy; and, *"medium"* otherwise. **Name in the database:** *redsoc_gradcare* **Source:** CONEVAL (2020)
-Social networks for entrepreneurship:	Degree of perception of the woman on the easiness to obtain support for entrepreneurship or economic purposes from social networks in three hypothetical circumstances: need of money, help to get a job, and collaboration to improve neighborhood conditions. **Type:** categorical. *"low"* if obtaining support from social networks in the majority of hypothetical situations is perceived by the woman as difficult or impossible; *"high"* if obtaining support from social networks in the majority of hypothetical situations is perceived by the woman as easy or very easy; and, *"medium"* otherwise. **Name in the database:** *redsoc_gradwork* **Source:** CONEVAL (2020)

Variable	Description
Relationship characteristics	
-Partner's weekly paid working hours:	Time in hours spent on paid work by the man per week. **Type:** continuous. **Name in the database:** *hor_1partner* **Source:** INEGI (2020)
-Partner's weekly leisure hours:	Time in hours spent on leisure by the man per week. **Type:** continuous. **Name in the database:** *hor_8partner* **Source:** INEGI (2020)
-Partner's age:	Age in years of the man. **Type:** continuous. **Name in the database:** *edadpartner* **Source:** INEGI (2020)
-Partner's income-to-poverty ratio:	The income-to-poverty ratio is calculated by dividing the labor income of the man by the poverty threshold. The official criteria for defining poverty indicates that a person is considered to be poor if their income is below the total cost of both the basic food basket and the non-food basket, which embraces transportation, education, health, entertainment, among others. For urban communities, the poverty threshold was set at MXN$ 3559.88 monthly *per capita*. **Type:** continuous. **Name in the database:** *ipovlabpartner* **Source:** INEGI (2020)
-Share of woman's labor income in total couple's labor income:	The share of woman's labor income in total couple's labor income calculated by dividing the labor income of the woman by the total couple's labor income (labor income of the woman plus the labor income of the man). **Type:** continuous. **Name in the database:** *ing_labPerc* **Source:** INEGI (2020)
-Partner's education level:	Degree of formal education level completed by the man. **Type:** categorical. *"low"* if the maximum completed level by the man is primary education; *"medium"* if the man has minimum secondary education and a maximum of high school; and, *"high"* if the man has completed at least a university degree. **Name in the database:** *nivelaprobpartner* **Source:** INEGI (2020)
-Partner's social networks for care:	Degree of perception of the man on the easiness to obtain support for care from social networks in three hypothetical circumstances: care due to illness, to be accompanied to a medical appointment, and child care assistance. **Type:** categorical. *"low"* if obtaining support from social networks in the majority of hypothetical situations is perceived by the man as difficult or impossible; *"high"* if obtaining support from social networks in the majority of hypothetical situations is perceived by the man as easy or very easy; and, *"medium"* otherwise. **Name in the database:** *redsoc_gradcarepartner* **Source:** CONEVAL (2020)
-Partner's social networks for entrepreneurship:	Degree of perception of the man on the easiness to obtain support for entrepreneurship or economic purposes from social networks in three hypothetical circumstances: need of money, help to get a job, and collaboration to improve neighborhood conditions. **Type:** categorical. *"low"* if obtaining support from social networks in the majority of hypothetical situations is perceived by the man as difficult or impossible;

Variable	Description
	"high" if obtaining support from social networks in the majority of hypothetical situations is perceived by the man as easy or very easy; and, *"medium"* otherwise. **Name in the database:** *redsoc_gradworkpartner* **Source:** CONEVAL (2020)
Household characteristics	
-Household members with income:	Share of household members having income. **Type:** continuous. **Name in the database:** *int_ing* **Source:** INEGI (2020)
-Children household members:	Share of household members under 12 years old. **Type:** continuous. **Name in the database:** *int_men* **Source:** INEGI (2020)
-Senior household members:	Share of household members aged 65 and more. **Type:** continuous. **Name in the database:** *int_65more* **Source:** INEGI (2020)
-Number of children:	Number of household members under 12 years old. **Type:** continuous. **Name in the database:** *child* **Source:** INEGI (2020)
-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* **Source:** INEGI (2020)
-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 (2020)
-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 (2020)

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 (2020)
-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 (2020)
-Type of household:	Type of household based on the number of members, and the relationship between them. **Type:** categorical. *"nuclear"* household consisting of the woman, and her partner; the woman, her partner, and their children; or, the woman, her partner, their children, and his/her parents. *"extended"* household consisting of the woman, her nuclear family, 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 woman, her nuclear family, 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 (2020)
-Children at home:	Children living in the same household. **Type:** categorical. *"yes"* if the at least one child is living in the same household; and, *"no"* otherwise. **Name in the database:** *child* **Source:** CONEVAL (2020)
-Parents at home:	Parents (or parents in-law) of the women living in the same household. **Type:** categorical. *"yes"* if the parents (or parents in-law) of the woman live in the same household; and, *"no"* otherwise. **Name in the database:** *parenthome* **Source:** CONEVAL (2020)

5.3 Data cleaning process for Section: Understanding gendered inequalities in time allocation to unpaid housework among partnered women and men in Mexico

After identifying the available relevant variables, the following analysis for each of the covariates is carried out:

- Plausibility. For this we inspect the information to discover potential incorrect coding or data errors, particularly in new covariates derived from existing variables such as social networks, education level, and age of the partner.
- Deletion of unusual observations. To do this, we inspect the following.
  - Households reporting more than one partner were deleted.
  - Households not reporting income from paid work were deleted (for the creation of variable "Share of woman's income in total household's labor income" the result would be undetermined).
- To ensure we have only complete cases in our data set, we delete all the observations with at least one missing value in the independent variables used.

5.4 Code for replicating the results in Section: Understanding gendered inequalities in time allocation to unpaid housework among partnered women and men in Mexico

```
######## Code for the replication of estimations in the Section 3.1:
Understanding gendered inequalities in ########
######## time allocation to unpaid housework among partnered
women and men in Mexico ########

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

### Database ###
load("Data␣and␣estimations␣code/database_timeuse.RData")
# Variables are already zero-centered

### Model ###
ftimeuse <- hor_6Dif ~ # housework time differential

  bols(intercept, intercept = FALSE) + # intercept

  bspatial(hor_1, hor_1partner, center = TRUE, df = 1,
  differences = 1, knots = 20) +

  bspatial(hor_8, hor_8partner, center = TRUE, df = 1,
  differences = 1, knots = 20) +

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

  bspatial(edad, edadpartner, center = TRUE, df = 1,
  differences = 1, knots = 20) +

  bspatial(ipovlab, ipovlabpartner, center = TRUE, df = 1,
  differences = 1, knots = 20) +

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

  bols(tarjeta, 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) +

#bols(nivelaproblow, intercept = FALSE, df = 1) +
bols(nivelaprobmedium, intercept = FALSE, df = 1) +
bols(nivelaprobhigh, intercept = FALSE, df = 1) +
#bols(edad, by = nivelaproblow, intercept = FALSE) +
#bbs(edad, by = nivelaproblow, center = TRUE,
df = 1, knots = 20) +
bols(edad, by = nivelaprobmedium, intercept = FALSE) +
bbs(edad, by = nivelaprobmedium, center = TRUE,
df = 1, knots = 20) +
bols(edad, by = nivelaprobhigh, intercept = FALSE) +
bbs(edad, by = nivelaprobhigh, center = TRUE,
df = 1, knots = 20) +

# bols(redsoc_gradcarelow, intercept = FALSE, df = 1) +
bols(redsoc_gradcaremedium, intercept = FALSE, df = 1) +
bols(redsoc_gradcarehigh, intercept = FALSE, df = 1) +

# bols(redsoc_gradworklow, intercept = FALSE, df = 1) +
bols(redsoc_gradworkmedium, intercept = FALSE, df = 1) +
bols(redsoc_gradworkhigh, intercept = FALSE, df = 1) +

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

#bols(nivelaprobpartnerlow, intercept = FALSE, df = 1) +
bols(nivelaprobpartnermedium, intercept = FALSE, df = 1) +
bols(nivelaprobpartnerhigh, intercept = FALSE, df = 1) +
bols(edadpartner, by = nivelaprobpartnermedium, intercept = FALSE) +
bbs(edadpartner, by = nivelaprobpartnermedium, center = TRUE,
df = 1, knots = 20) +
bols(edadpartner, by = nivelaprobpartnerhigh, intercept = FALSE) +
bbs(edadpartner, by = nivelaprobpartnerhigh, center = TRUE,
df = 1, knots = 20) +

# bols(redsoc_gradcarepartnerlow, intercept = FALSE, df = 1) +
bols(redsoc_gradcarepartnermedium, intercept = FALSE, df = 1) +
bols(redsoc_gradcarepartnerhigh, intercept = FALSE, df = 1) +

# bols(redsoc_gradworkpartnerlow, intercept = FALSE, df = 1) +
bols(redsoc_gradworkpartnermedium, intercept = FALSE, df = 1) +
bols(redsoc_gradworkpartnerhigh, intercept = FALSE, df = 1) +
```

```
  # bols(clase_hognuclear, intercept = FALSE, df = 1) +
  bols(clase_hogextended, intercept = FALSE, df = 1) +
  bols(clase_hogother, intercept = FALSE, df = 1) +

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

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

  bols(int_65more, intercept = FALSE) +
  bbs(int_65more, center = TRUE, df = 1, knots = 20) +

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

  bols(edad, by = children, intercept = FALSE) +
  bbs(edad, by = children, center = TRUE, df = 1, knots = 20) +

  bols(edad, by = parenthome, intercept = FALSE) +
  bbs(edad, by = parenthome, center = TRUE, df = 1, knots = 20)

### Three-step strategy ###
## Functional gradient descent boosting
modTimeDiff <- gamboost(ftimeuse,
                        data = urbanwomen,
                        control = boost_control(mstop = 1000,
                                                nu = 0.25,
                                                trace = TRUE,
                                                stopintern = TRUE),
                        offset = weighted.mean(urbanwomen$hor_6Dif,
                        w = urbanwomen$factor),
                        weights = urbanwomen$factor,
                        family = Gaussian())

# Cross-validation
set.seed(1806)
cvwomenDif <- cvrisk(modTimeDiff,
                     folds = cv(model.weights(modTimeDiff),
                                type = "subsampling"),
                     grid = 1:3000,
                     papply = mclapply,
                     mc.cores = parallel::detectCores())
stopwomenDif <- mstop(cvwomenDif)
modTimeDiff[stopwomenDif]
summary(modTimeDiff)

## Stability selection
p <- length(names(coef(modTimeDiff, which = "")))
```

```
stabsel_parameters(p = p, q = 10, 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)

stabTimewomenDif <- stabsel(modTimeDiff,
                            cutoff = 0.80,
                            q = 10,
                            sampling.type = "SS",
                            mc.cores = parallel::detectCores())

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

save(modTimeDiff, ciTimewomenDif, stopwomenDif, cvwomenDif,
stabTimewomenDif,
file = "Data␣and␣estimations␣code/results_timeuse.RData")
```

5.5 Metadata for the data used in Section: Emotional IPV against women and girls with children in Mexican households

The following are the metadata describing the attributes needed to use and understand the data utilized in the analysis developed in Section 3.2. The full data set can be found in the file called "database_ipv.RData" and is freely available from Figshare at https://doi.org/10.6084/m9.figshare.21183271.

Variable	Description
Individual-level covariates	
-Woman's age:	Age in years of the woman. **Type:** continuous. **Name in the database:** *EDAD* **Source:** INEGI (2016c)
-Woman's income:	Woman's reported monthly earned income, in Mexican Pesos. **Type:** continuous. **Name in the database:** *ingm_muj* **Source:** INEGI (2016c)
-Woman's age at first childbirth:	Age in years of the woman at first childbirth. **Type:** continuous. **Name in the database:** *eda_hij* **Source:** INEGI (2016c)
-Woman's age at her first sexual intercourse:	Age in years of the woman at her first sexual intercourse. **Type:** continuous. **Name in the database:** *eda_sex* **Source:** INEGI (2016c)
-Indigenous origin:	Indigenous self-identification of the woman. **Type:** categorical. *"yes"* if the woman self identifies as indigenous; and, *"no"* otherwise. **Name in the database:** *indigena* **Source:** INEGI (2016c)
-Education level:	Degree of formal education level completed by the woman. **Type:** categorical. *"low"* if the maximum completed level by the woman is primary education; *"medium"* if the woman has minimum secondary education and a maximum of high school; and, *"high"* if the woman has completed at least a university degree. **Name in the database:** *niv_ed* **Source:** INEGI (2016c)
-Woman's consent to first sexual intercourse:	Consent to first sexual intercourse. **Type:** categorical. *"yes"* if the woman consented to her first sexual experience; and, *"no"* otherwise. **Name in the database:** *con_sex* **Source:** INEGI (2016c)

Variable	Description
-Pro-gender equality attitude:	Opinion on gender roles in taking care of children, income, responsibility for providing money to the household, responsibility for carrying out housework, women's right to go out at night, if men should have better jobs than women, if women with kids and working neglect their responsibilities as mother, sex in marriage and type of clothes women should use. **Type:** categorical. *"low"* if the woman has a pro-gender equality opinion in 2 or less questions; *"medium"* if the woman has a pro-gender equality opinion in more than 2 but less than 7 questions; and, *"high"* if the woman has a pro-gender equality opinion in at least 7 questions. **Name in the database:** *feminist_grad* **Source:** INEGI (2016c)
Relationship-level covariates	
-Woman's age at marriage or at cohabitation:	Age in years of the woman aat marriage with current husband or at cohabiting with current partner. **Type:** continuous. **Name in the database:** *eda_mat* **Source:** INEGI (2016c)
-Partners's age:	Age in years of the partner. **Type:** continuous. **Name in the database:** *eda_par2* **Source:** INEGI (2016c)
-Partner's income:	Partner's reported monthly earned income, in Mexican Pesos. **Type:** continuous. **Name in the database:** *ingm_par* **Source:** INEGI (2016c)
-Overcrowding:	Average number of household members per room in the dwelling. **Type:** continuous. **Name in the database:** *hacin* **Source:** INEGI (2016c)
-Woman's consent to marriage or cohabitation:	Consent for marriage with current husband or for cohabiting with current partner. **Type:** categorical. *"yes"* if the woman consented to marriage or cohabitation; and, *"no"* otherwise. **Name in the database:** *con_mat* **Source:** INEGI (2016c)
-Division of housework among household members:	Division of housework among household members. **Type:** categorical. *"only women"* if housework at home is carried out only by women members; *"both"* if both women and men take part in housework; and, *"only men"* if housework is carried out only by men members. **Name in the database:** *act_dist* **Source:** INEGI (2016c)
-Woman's level of autonomy within the relationship to make decisions about her sexual life:	Perception of the woman regarding her level of freedom and autonomy within the relationship to take decisions about her sexual life, particularly about when to have sex, whether and who should use contraceptive methods, whether and when to have children. **Type:** categorical. *"low"* if the woman expressed having none or less decision making power than her husband or partner; *"medium"* if the woman expressed having the same decision making power than her husband or partner; and, *"high"* if the woman expressed having all or more decision making power than her husband or partner.

Variable	Description
	Name in the database: *lib_sex_grad* **Source:** INEGI (2016c)
-Woman's level of autonomy within the relationship to make decisions about her professional life and use of economic resources:	Perception of the woman regarding her level of freedom and autonomy within the relationship to take decisions about her professional life and use of economic resources, particularly about whether working or studying, use of her money, and buying things for her. **Type:** categorical. *"low"* if the woman expressed having none or less decision making power than her husband or partner; *"medium"* if the woman expressed having the same decision making power than her husband or partner; and, *"high"* if the woman expressed having all or more decision making power than her husband or partner. **Name in the database:** *lib_eco_grad* **Source:** INEGI (2016c)
-Woman's level of autonomy within the relationship to make decisions about her participation in social and political activities:	Perception of the woman regarding her level of freedom and autonomy within the relationship to take decisions about her participation in social and political activities, particularly about whether going out of home, and participating in social and political life of the community. **Type:** categorical. *"low"* if the woman expressed having none or less decision making power than her husband or partner; *"medium"* if the woman expressed having the same decision making power than her husband or partner; and, *"high"* if the woman expressed having all or more decision making power than her husband or partner. **Name in the database:** *lib_soc_grad* **Source:** INEGI (2016c)
-Social networks:	Perception of the woman whether she could get support from social networks in some hypothetical situations, including if the woman needed child care assistance, carrying out a task or work, care due to illness, talk about her problems and worries, have advice or guidance when she has a problem with her husband or partner, and have support when she is in difficult economic times. **Type:** categorical. *"low"* if the woman considers she could get support from social networks in 1 or less situations; *"medium"* if the woman considers she could get support from social networks in more than 1 but less than 5 situations; and, *"high"* if the woman considers she could get support from social networks in at least 5 situations. **Name in the database:** *redsoc_grad* **Source:** INEGI (2016c)
-Level of social interaction reported by the woman:	Level of social interaction reported by the woman, including going out with friends, talking with neighbors, meeting with family members, attending religious events, participating in organizations, and practicing team sports. **Type:** categorical. *"low"* if the woman states she carries out 1 or less situations; *"medium"* if the woman states she carries out at least more than 1 but less than 5 situations; and, *"high"* if the woman states she carries out at least 5 situations. **Name in the database:** *rout_grad* **Source:** INEGI (2016c)
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)
-Type of community:	Type of community according to their number of inhabitants. **Type:** categorical. *"rural"* if the community has less than 2,500 inhabitants; *"low urban"* if the community has between 2,500 and 14,999 inhabitants; *"medium urban"* if the community has between 15,000 and 99,999 inhabitants; and, *"high urban"* if the community has 100,000 inhabitants or more inhabitants. **Name in the database:** *Type_com* **Source:** CONAPO (2016)
-Women homicide rate:	Yearly average number of women homicides from 2013 to 2017 per 100,000 women considering the 2015 population by Municipality. **Type:** continuous. **Name in the database:** *fhr15* **Source:** INEGI (2022)
-Men homicide rate:	Yearly average number of men homicides from 2013 to 2017 per 100,000 men considering the 2015 population by Municipality. **Type:** continuous. **Name in the database:** *mhr15* **Source:** INEGI (2022)
-Total homicide rate:	Yearly average number of total homicides from 2013 to 2017 per 100,000 inhabitants considering the 2015 population by Municipality. **Type:** continuous. **Name in the database:** *ghr15* **Source:** INEGI (2022)
-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)
-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.

Variable	Description
	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)
-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'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'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. **Name in the database:** *ParPolF* **Source:** INEGI (2015a)
Societal-level covariates	
-Common crimes against women:	Prevalence rate of common crimes against women aged 18 or more per 100,000 women in 2015. **Type:** continuous. **Name in the database:** *FemPrev* **Source:** INEGI (2016b)
-Common crimes against men:	Prevalence rate of common crimes against men aged 18 or more per 100,000 men in 2015. **Type:** continuous. **Name in the database:** *MasPrev* **Source:** INEGI (2016b)
-Dark figure of common crimes against women:	-Share of common crimes against women aged 18 or more not reported to or not registered by the authorities in 2015. **Type:** continuous.

Variable	Description
	Name in the database: *FemNoDen* **Source:** INEGI (2016b)
-Dark figure of common crimes against men:	-Share of common crimes against men aged 18 or more not reported to or not registered by the authorities in 2015. **Type:** continuous. **Name in the database:** *MasNoDen* **Source:** INEGI (2016b)
-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)

5.6 Data integration process for Section: Examining gender inequalities in factors associated with income poverty in Mexican rural households

The process of bringing the ENIGH information and the independent variables from disparate sources together to generate a unified view to be modeled is described in the following lines:

- First, from the ENIGH microdata we select the information related to rural households. Based on existing theories and research, we then subset these data to select only the information about the dependent and independent variables at the individual and household level. Each of these observations 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 (2016a).

- Data at the community level are taken from the Intercensal Population Survey, CONAPO, UNDP,CONEVAL, CNGMD, and municipal geographic coordinates by INEGI. Utilizing the municipality unique identifier assigned by INEGI CVE_MUN as a common variable among the data sets, we first join all the data at the municipal level from these sources, and then we combine them with the ENIGH microdata. This results in a database with a two-dimensional tree-like hierarchical structure, in which the individual and household observations of the ENIGH microdata (first dimension) are linked to the data at the community level (second dimension).

- Finally, the estimations at the state level from the ENCIG and the ENVIPE, which contain the state unique identifier assigned by INEGI, CVE_ENT, are combined with the data 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 linked to the information at the municipal level (second dimension), and these, in turn, to the state level estimations (third dimension).

5.7 Data integration process for Section: Emotional IPV against women and girls with children in Mexican households

The process of integrating the ENDIREH with the other ten data sources consists of three steps:

- From the ENDIREH microdata we select the information related to the questionnaire applied to married or cohabitation women. The observations in these microdata correspond to individual answers given by the respondents to the ENDIREH 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 (2016c).

- Estimations at the municipal level from the Intercensal Population Survey, CONAPO, UNDP,CONEVAL, homicide records, CNGMD, ENCIG, and the geographic information also contain the municipality unique identifier assigned by INEGI, CVE_MUN. Using this CVE_MUN as a common variable among the data sets, we first merge all the data at the municipal level from these sources, before merging them with the ENDIREH microdata. This results in a database with a two-dimensional tree-like hierarchical structure, in which the individual observations of the ENDIREH microdata (first dimension) are connected to the estimations at the municipal level (second dimension).

- Finally, the estimations at the state level from the ENCIG and the ENVIPE, 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 ENDIREH 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).

5.8 Data cleaning process for Section: Emotional IPV against women and girls with children in Mexican households

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

- Plausibility. This process consists of inspecting the data to discover potential incorrect coding or data errors, particularly in new covariates derived from existing variables. The three situations analyzed are:
  - Women's age at first sexual intercourse cannot be greater than women's age at the time of being surveyed.
  - Women's age at first marriage (or at cohabitation) cannot be greater than women's age at the time of being surveyed.
  - Women's age at first childbirth cannot be greater than women's age at the time of being surveyed. No implausible values were found.
- Outlier detection. To prevent a few unusual observations from influencing the results, we identify the extreme values and exclude them from the final data. To do this, we create boxplots for the continuous variables.
- To ensure we have only complete cases in our data set, we delete all the observations with at least one missing value in the independent variables used.

5.9 Code for replicating the results in Section: Emotional IPV against women and girls with children in Mexican households

```
######## Code for the replication of estimations in the Section 3.2:
######## Emotional ipv against women and girls ########
######## with children in Mexican households ########

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

### Database ###
load("Data␣and␣estimations␣code/database_ipv.RData")
# Variables are already zero-centered

### Model ###
vio_emo_año <- vio_emo_año ~

  bols(intercept, intercept = FALSE) +

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

  bols(EDAD, by = indigena, intercept = FALSE) +
  bbs(EDAD, by = indigena, knots = 20, df = 1, center = TRUE) +

  bols(EDAD, by = niv_edmedium, intercept = FALSE) +
  bbs(EDAD, by = niv_edmedium, knots = 20, df = 1, center = TRUE) +
  bols(EDAD, by = niv_edhigh, intercept = FALSE) +
  bbs(EDAD, by = niv_edhigh, knots = 20, df = 1, center = TRUE) +

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

  bspatial(eda_hij, EDAD, center = TRUE, differences = 1,
  knots = 20, df = 1) +

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

  bols(eda_sex, by = con_sex, intercept = FALSE) +
  bbs(eda_sex, by = con_sex, knots = 20, df = 1, center = TRUE) +

  bspatial(eda_sex, EDAD, center = TRUE, differences = 1,
  knots = 20, df = 1) +
```

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

bols(eda_mat, by = mot_mat, intercept = FALSE) +
bbs(eda_mat, by = mot_mat, knots = 20, df = 1, center = TRUE) +

bspatial(eda_mat, EDAD, center = TRUE, differences = 1,
knots = 20, df = 1) +

bols(eda_par2, intercept = FALSE) +
bbs(eda_par2, knots = 20, df = 1, center = TRUE) +

bspatial(eda_par2, EDAD, center = TRUE, differences = 1,
knots = 20, df = 1) +

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

bols(act_distboth, intercept = FALSE, df = 1) +
bols(act_distmales, intercept = FALSE, df = 1) +

bols(feminist_gradmedium, intercept = FALSE, df = 1) +
bols(feminist_gradhigh, intercept = FALSE, df = 1) +

bols(lib_sex_gradmedium, intercept = FALSE, df = 1) +
bols(lib_sex_gradhigh, intercept = FALSE, df = 1) +

bols(lib_eco_gradmedium, intercept = FALSE, df = 1) +
bols(lib_eco_gradhigh, intercept = FALSE, df = 1) +

bols(lib_soc_gradmedium, intercept = FALSE, df = 1) +
bols(lib_soc_gradhigh, intercept = FALSE, df = 1) +

bols(redsoc_gradmedium, intercept = FALSE, df = 1) +
bols(redsoc_gradhigh, intercept = FALSE, df = 1) +

bols(rout_gradmedium, intercept = FALSE, df = 1) +
bols(rout_gradhigh, intercept = FALSE, df = 1) +

bspatial(ingm_muj, ingm_par, center = TRUE, differences = 1,
knots = 20, df = 1) +

brandom(cvegeo, df = 1) +

bols(mhr15, intercept = FALSE) +
bbs(mhr15, knots = 20, df = 1, center = TRUE) +

bols(fhr15, intercept = FALSE) +
```

```
bbs(fhr15, knots = 20, df = 1, center = TRUE) +

bols(ghr15, intercept = FALSE) +
bbs(ghr15, knots = 20, df = 1, center = TRUE) +

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

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

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

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

bols(idh2015, intercept = FALSE) +
bbs(idh2015, knots = 20, df = 1, center = TRUE) +

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

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

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

bols(Marg15low, intercept = FALSE, df = 1) +
bols(Marg15medium, intercept = FALSE, df = 1) +
bols(Marg15high, intercept = FALSE, df = 1) +
bols(Marg15very_high, intercept = FALSE, df = 1) +

bols(Type_comlow_urban, intercept = FALSE, df = 1) +
bols(Type_commedium_urban, intercept = FALSE, df = 1) +
bols(Type_comhigh_urban, intercept = FALSE, df = 1) +

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

brandom(cveent, df = 1) +

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

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

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

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

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

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

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

### Three-step strategy ###
## Functional gradient descent boosting
modelemoipv <- gamboost(vio_emo_año,
                        data = vawgdbOutcC,
                        control = boost_control(mstop = 2000,
                                                nu = 0.5,
                                                trace = TRUE,
                                                stopintern = TRUE),
                        weights = vawgdbOutcC$FAC_MUJ,
                        offset = pnorm(weighted.mean(
                        x = as.numeric(as.character(
                        vawgdbOutcC[, "vio_emo_año"])),
                        w = vawgdbOutcC$FAC_MUJ))-0.5,
                        family = Binomial(link = "probit"))

# Cross-validation
set.seed(1806)
cvemoipv <- cvrisk(modelemoipv,
                   folds = cv(model.weights(modelemoipv),
                              type = "subsampling"),
                   grid = 1:10000,
                   papply = mclapply,
                   mc.cores = parallel::detectCores())
stopemoipv <- mstop(cvemoipv)
modelemoipv[stopemoipv]

## Stability selection
p <- length(names(coef(modelemoipv, which = "")))
stabsel_conf <- stabsel_parameters(p = p,
                                   q = 20,
                                   cutoff = 0.8)
# Stability selection with unimodality assumption

# Cutoff: 0.8; q: 20; PFER (*):  3.74
```

```
# (*) or expected number of low selection probability variables
# PFER (specified upper bound):  3.743316
# PFER corresponds to signif. level 0.0425
# (without multiplicity adjustment)
stabselemoipv <- stabsel(modelemoipv,
                         q = 20,
                         cutoff = 0.8,
                         sampling.type = "SS",
                         mc.cores = parallel::detectCores())

## Pointwise bootstrap confidence intervals
confintemoipv <- confint(modelemoipv, B = 1000,
                         level = 0.95, B.mstop = 0,
                         papply = mclapply,
                         cvrisk_options = list(mc.cores = 25))

save(confintemoipv, stabselemoipv, modelemoipv, stopemoipv, cvemoipv,
file = "estimation_ipv.RData")
```

5.10 Metadata for the data used in Section: Examining gender inequalities in factors associated with income poverty in Mexican rural households

The following are the metadata describing the attributes needed to use and understand the data utilized in the analysis developed in Section 3.3. The full data set can be found in the file called "database_poverty.RData" and is freely available from Figshare at https://doi.org/10.6084/m9.figshare.21183271.

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;

Variable	Description
	"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* **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.

Variable	Description
	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)
-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*

Variable	Description
	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	
-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:** CENAPRED (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)

Variable	Description
-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. **Name in the database:** *ParPolF* **Source:** INEGI (2015a)
-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*

Variable	Description
	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)
-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 (2016c)
-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 (2016c)

Variable	Description
-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 (2016c)
-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. **Name in the database:** *TPrevRel12Mes* **Source:** INEGI (2016c)
-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 (2016c)

5.11 Data integration process for Section: Examining gender inequalities in factors associated with income poverty in Mexican rural households

The process of bringing the ENIGH information and the independent variables from disparate sources together to generate a unified view to be modeled is described in the following lines:

- First, from the ENIGH microdata we select the information related to rural households. Based on existing theories and research, we then subset these data to select only the information about the dependent and independent variables at the individual and household level. Each of these observations 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 (2016a).

- Data at the community level are taken from the Intercensal Population Survey, CONAPO, UNDP,CONEVAL, CNGMD, and municipal geographic coordinates by INEGI. Utilizing the municipality unique identifier assigned by INEGI CVE_MUN as a common variable among the data sets, we first join all the data at the municipal level from these sources, and then we combine them with the ENIGH microdata. This results in a database with a two-dimensional tree-like hierarchical structure, in which the individual and household observations of the ENIGH microdata (first dimension) are linked to the data at the community level (second dimension).

- Finally, the estimations at the state level from the ENCIG and the ENVIPE, which contain the state unique identifier assigned by INEGI, CVE_ENT, are combined with the data 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 linked to the information at the municipal level (second dimension), and these, in turn, to the state level estimations (third dimension).

5.12 Data cleaning process for Section: Examining gender inequalities in factors associated with income poverty in Mexican rural households

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

- Plausibility. This process consists of inspecting the data to discover potential incorrect coding or data errors.
- Outlier detection. To prevent a few unusual observations from influencing the results, we identify the extreme values and exclude them from the final data. To do this, we create boxplots for the continuous variables.
- To ensure we have only complete cases in our data set, we delete all the observations with at least one missing value in the independent variables used.

5.13 Code for replicating the results in Section: Examining gender inequalities in factors associated with income poverty in Mexican rural households

```
######## Code for the replication of estimations in the Section 3.3:
Examining gender inequalities in factors ########
######## associated with income poverty in Mexican rural households
########

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

### Database ###
load("Data␣and␣estimations␣code/database_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 ###
## Functional gradient descent boosting
modelmext <- gamboost(fPoverty,
                      data = mictpcRu,
                      control = boost_control(mstop = 5000,
                                              nu = 0.50,
                                              trace = TRUE,
                                              stopintern = TRUE),
                      weights = mictpcRu$factor,
                      family = QuantReg(
                      tau = ecdf(mictpcRu$ipov)(933.20/1715.57)),
                      offset = quantile(x = mictpcRu$ipov,
                      prob = ecdf(mictpcRu$ipov)(933.20/1715.57)))

# 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 = fictpcRu,
                      control = boost_control(mstop = 5000,
                                              nu = 0.50,
                                              trace = TRUE,
                                              stopintern = TRUE),
                      weights = fictpcRu$factor,
                      family = QuantReg(
                      tau = ecdf(fictpcRu$ipov)(933.20/1715.57)),
                      offset = quantile(x = fictpcRu$ipov,
                      prob = ecdf(fictpcRu$ipov)(933.20/1715.57)))

# 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 = mictpcRu,
                      control = boost_control(mstop = 5000,
                                              nu = 0.50,
                                              trace = TRUE,
                                              stopintern = TRUE),
                      weights = mictpcRu$factor,
                      family = QuantReg(
                      tau = ecdf(mictpcRu$ipov)(1715.57/1715.57)),
                      offset = quantile(x = mictpcRu$ipov,
                      prob = ecdf(mictpcRu$ipov)(1715.57/1715.57)))

# 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 = fictpcRu,
                      control = boost_control(mstop = 5000,
                                              nu = 0.50,
                                              trace = TRUE,
                                              stopintern = TRUE),
                      weights = fictpcRu$factor,
                      family = QuantReg(
                      tau = ecdf(fictpcRu$ipov)(1715.57/1715.57)),
                      offset = quantile(x = fictpcRu$ipov,
                      prob = ecdf(fictpcRu$ipov)(1715.57/1715.57)))

# 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 selection
p <- length(names(coef(modelfpov, 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)

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")
```

Bibliography

Abramsky, T., Lees, S., Stöckl, H., Harvey, S., Kapinga, I., Ranganathan, M., Mshana, G., & Kapiga, S. (2019). Women's income and risk of intimate partner violence: Secondary findings from the maisha cluster randomised trial in north-western tanzania. *BMC public health*, *19*(1), 1108. https://doi.org/10.1186/s12889-019-7454-1

Achjar, N., & Panennungi, M. A. (2009). The impact of rural infrastructure development on poverty reduction in indonesia. *Economics and Finance in Indonesia*, *57*, 339–348.

Ackerson, L. K., & Subramanian, S. V. (2008). State gender inequality, socioeconomic status and intimate partner violence (ipv) in india: A multilevel analysis. *Australian Journal of Social Issues*, *43*(1), 81–102. https://doi.org/10.1002/j.1839-4655.2008.tb00091.x

Adeyonu, A. G., & Oni, O. A. (2014). Gender time allocation and farming households' poverty in rural nigeria. *World Journal of Agricultural Sciences*, *2*(5), 123–136. https://eprints.lmu.edu.ng/628/

Ahmadabadi, Z., Najman, J. M., Williams, G. M., & Clavarino, A. M. (2020). Income, gender, and forms of intimate partner violence. *Journal of interpersonal violence*, *35*(23-24), 5500–5525. https://doi.org/10.1177/0886260517719541

Alon, T., Coskun, S., Doepke, M., Koll, D., & Tertilt, M. (2021). From mancession to shecession: Women's employment in regular and pandemic recessions. *National Bureau of Economic Research*. https://doi.org/10.3386/w28632

Alvarado-Zaldívar, G., Moysén, J. S., Estrada-Martínez, S., & Terrones-González, A. (1998). Prevalencia de violencia doméstica en la ciudad de durango. *Salud Pública de México*, *40*(6), 481–486. https://doi.org/10.1590/S0036-36341998000600004

Álvarez, B., & Miles, D. (2004). Husbands' housework time: Does wives' paid employment make a difference? *Investigaciones Economicas*, *30*(1), 5–31. https://econpapers.repec.org/paper/vigwpaper/0402.htm

Anderssen, N., & Wold, B. (1992). Parental and peer influences on leisure-time physical activity in young adolescents. *Research Quarterly for Exercise and Sport*, *63*(4), 341–348. https://doi.org/10.1080/02701367.1992.10608754

Avila-Burgos, L., Valdez-Santiago, R., Híjar, M., Del Rio-Zolezzi, A., Rojas-Martínez, R., & Medina-Solís, C. E. (2009). Factors associated with severity of intimate partner abuse in mexico: Results of the first national survey of violence against women. *Canadian Journal of Public Health*, *100*(6), 436–441. http://www.jstor.org/stable/41995320

Baez, J. E., Kronick, D., & Mason, A. D. (2013). Rural households in a changing climate. *The World Bank Research Observer*, *28*(2), 267–289. https://doi.org/10.1093/wbro/lks008

Barbier, E. B. (2012). Corruption, poverty and tropical land use. *Journal of Sustainable Forestry*, *31*(4-5), 319–339. https://doi.org/10.1080/10549811.2011.588455

Begoña Álvarez, D. M. (2006). Husbands' housework time: Does wives' paid employment make a diference? *Investigaciones Económicas*. https://www.redalyc.org/articulo.oa?id=17330101

Belloni, A., Chernozhukov, V., & Hansen, C. (2014). High-dimensional methods and inference on structural and treatment effects. *Journal of Economic Perspectives*, *28*(2), 29–50. https://doi.org/10.1257/jep.28.2.29

Bianchi, S. M., Milkie, M. A., Sayer, L. C., & Robinson, J. P. (2000). Is anyone doing the housework? trends in the gender division of household labor. *Social Forces*, *79*(1), 191. https://doi.org/10.2307/2675569

Bloch, F., & Rao, V. (2002). Terror as a bargaining instrument: A case study of dowry violence in rural india. *The American Economic Review*, *92*(4), 1029–1043.

Bogale, A., Hagedorn, K., & Korf, B. (2005). Determinants of poverty in rural ethiopia. https://doi.org/10.5167/UZH-64170

Bühlmann, P. (2006). Boosting for high-dimensional linear models. *The Annals of Statistics*, *34*(2), 559–583. https://doi.org/10.1214/009053606000000092

Bühlmann, P., & Hothorn, T. (2007). Boosting Algorithms: Regularization, Prediction and Model Fitting. *Statistical Science*, *22*(4), 477–505. https://doi.org/10.1214/07-STS242

Caetano, R., Field, C., Ramisetty-Mikler, S., & McGrath, C. (2005). The 5-year course of intimate partner violence among white, black, and hispanic couples in the united states. *Journal of interpersonal violence*, *20*(9), 1039–1057. https://doi.org/10.1177/0886260505277783

Caetano, R., Schafer, J., & Cunradi, C. B. (2001). Alcohol-related intimate partner violence among white, black, and hispanic couples in the united states. *Alcohol research & health : the journal of the National Institute on Alcohol Abuse and Alcoholism*, *25*(1), 58–65. https://doi.org/Study

Cameron, A., & Tedds, L. M. (2021). *Gender-based violence, economic security, and the potential of basic income: A discussion paper*. https://mpra.ub.uni-muenchen.de/107478/

Casique, I., & Castro, R. (2014). *Expresiones y contextos de la violencia contra las mujeres en méxico: Resultados de la endireh 2011 en comparación con sus versiones previas 2003 y 2006*.

Castro, R., & Casique, I. (2008). *Violencia de género en las parejas mexicanas. análisis de resultados de la encuesta nacional sobre la dinámica de las relaciones en los hogares, 2006*.

Castro, R., & Casique, I. (2009). *Violencia de pareja contra las mujeres en méxico: Una comparación entre encuestas recientes* (Vol. 87).

Castro, R., Riquer, F., & Medina, M. E. (2006). *Violencia de género en las parejas mexicanas. resultados de la encuesta nacional sobre la dinámica de las relaciones en los hogares 2003*.

CENAPRED. (2020). Sistema de consulta de declaratorias. http://www.atlasnacionalderiesgos.gob.mx/apps/Declaratorias/

Chaurasia, H., Debnath, P., Srivastava, S., & Purkayastha, N. (2021). Is socioeconomic inequality boosting intimate partner violence in india? an overview of the national family health survey, 2005–2006 and 2015–2016. *Global Social Welfare*, 1–15. https://doi.org/10.1007/s40609-021-00215-6

Christensen, R. (2019). *Advanced linear modeling [electronic resource]: Statistical learning and dependent data / ronald christensen* (3rd ed.). Springer.

CONAPO (Ed.). (2016). *La situación demográfica de méxico 2016: Mujeres jefas de hogar y algunas características de los hogares que dirigen. una visión sociodemográfica*.

CONEVAL. (2018). Medición de la pobreza 2008 - 2018: Programas de cálculo. https://www.coneval.org.mx/Medicion/MP/Paginas/Programas_BD_08_10_12_14_16_18.aspx

CONEVAL. (2019). Metodología para la medición multidimensional de la pobreza en méxico (3rd ed.).

CONEVAL. (2020). Medición de la pobreza: Evolución de las líneas de pobreza por ingresos. https://www.coneval.org.mx

CONEVAL. (2021). Cohesión social. https://www.coneval.org.mx/Medicion/Paginas/Cohesion_Social.aspx

Cuevas Hernández, A. J. (2010). Jefas de familia sin pareja: Estigma social y autopercepción. *Estudios Sociológicos*, *28*(84), 753–789.

Datta Gupta, N., & Stratton, L. S. (2010). Examining the impact of alternative power measures on individual time use in american and danish couple households. *Review of Economics of the Household*, *8*(3), 325–343. https://doi.org/10.1007/s11150-009-9073-6

de Janvry, A., & Sadoulet, E. (2000). Rural poverty in latin america: Determinants and exit paths. *Food Policy*, *25*(4), 389–409.

Deyshappriya, N. R., & Minuwanthi, R. (2020). Determinants of poverty: Is age non-linearly related with poverty? evidence from sri lanka. *International Journal of Asian Social Science*, *10*(4), 181–192. https://doi.org/10.18488/journal.1.2020.104.181.192

Dias, N. G., Fraga, S., Soares, J., Hatzidimitriadou, E., Ioannidi-Kapolou, E., Lindert, J., Sundin, Ö., Toth, O., Barros, H., & Ribeiro, A. I. (2020). Contextual determinants of intimate partner violence: A multi-level analysis in six european cities. *International Journal of Public Health*, *65*(9), 1669–1679. https://doi.org/10.1007/s00038-020-01516-x

Eilers, P. H. C., & Marx, B. D. (1996). Flexible smoothing with b -splines and penalties. *Statistical Science*, *11*(2), 89–121. https://doi.org/10.1214/ss/1038425655

Espino, I., Hermeto, A., & Luz, L. (2020). Gender differences in time allocation to paid and unpaid work: Evidence from urban guatemala, 2000-2014. *MPRA Paper No. 106477*. https://mpra.ub.uni-muenchen.de/106477/

Esquivel-Santoveña, E. E., Hernández, R. R., Viveros, N. C., Orozco, F. L., & van Barneveld, H. O. (2020). Physical intimate partner violence and controlling behavior in mexican university students and their attitudes toward social limits. *Journal of interpersonal violence*, *35*(1-2), 403–425. https://doi.org/10.1177/0886260516681879

Fahrmeir, L., Kneib, T., Lang, S., & Marx, B. (2013). *Regression: Models, methods and applications*. Springer.

Fang, L., & McDaniel, C. (2017). Home hours in the united states and europe. *The B.E. Journal of Macroeconomics*, *17*(1). https://doi.org/10.1515/bejm-2015-0031

Fenske, N., Kneib, T., & Hothorn, T. (2011). Identifying risk factors for severe childhood malnutrition by boosting additive quantile regression. *Journal of the American Statistical Association*, *106*(494), 494–510. https://doi.org/10.1198/jasa.2011.ap09272

Ferrant, G., Pesando, L. M., & Nowacka, K. (2014). *Unpaid care work: The missing link in the analysis of gender gaps in labour outcomes*.

Field, E., Pande, R., Rigol, N., Schaner, S., & Moore, C. (2016). *On her account: Can strengthening women's financial control boost female labor supply?* (Vol. Working Paper No. 32).

Frías, S. M. (2017). Challenging the representation of intimate partner violence in mexico: Unidirectional, mutual violence and the role of male control. *Partner Abuse*, *8*(2), 146–167. https://doi.org/10.1891/1946-6560.8.2.146

Friedman, J. (2001). Greedy function approximation: A gradient boosting machine. *The Annals of Statistics*, *29*(5), 1189–1232. https://doi.org/10.2307/2699986

Friedman, J., Hastie, T., & Tibshirani, R. (2000). Additive logistic regression: A statistical view of boosting (with discussion and a rejoinder by the authors). *The Annals of Statistics*, *28*(2), 337–407. https://doi.org/10.1214/aos/1016218223

García-Ramos, A. (2021). Divorce laws and intimate partner violence: Evidence from mexico. *Journal of Development Economics*, *150*, 102623. https://doi.org/10.1016/j.jdeveco.2020.102623

Gashaw, B. T., Schei, B., & Magnus, J. H. (2018). Social ecological factors and intimate partner violence in pregnancy. *PloS one*, *13*(3), e0194681. https://doi.org/10.1371/journal.pone.0194681

Gillanders, R., & van der Werff, L. (2020). Corruption experiences and attitudes to political, interpersonal, and domestic violence. *MPRA Paper*, (99949).

Gimenez-Nadal, J. I., & Molina, J. A. (2020). *The gender gap in time allocation in europe*. Bonn, Germany : IZA - Institute of Labor Economics. http://ftp.iza.org/dp13461.pdfhttps://www.iza.org/publications/

dp/13461/the-gender-gap-in-time-allocation-in-europehttp://hdl.handle.net/10419/223903

Giurge, L. M., Whillans, A. V., & Yemiscigil, A. (2021). A multicountry perspective on gender differences in time use during covid-19. *Proceedings of the National Academy of Sciences of the United States of America*, *118*(12). https://doi.org/10.1073/pnas.2018494118

González, L., & Rodríguez-Planas, N. (2020). Gender norms and intimate partner violence. *Journal of Economic Behavior & Organization*, *178*, 223–248. https://doi.org/10.1016/j.jebo.2020.07.024

Gupta, S. (2006). Her money, her time: Women's earnings and their housework hours. *Social Science Research*, *35*(4), 975–999. https://doi.org/10.1016/j.ssresearch.2005.07.003

Hamilton, A., & Svensson, J. (2017). The vicious circle of poverty, poor public service provision, and state legitimacy in sudan. In A. Hamilton & C. Hammer (Eds.), *Data-driven decision making in fragile contexts: Evidence from sudan* (pp. 107–117). The World Bank. https://doi.org/10.1596/978-1-4648-1064-0{\textunderscore}ch6

Harrell Jr., F. E. (2015). *Regression modeling strategies: With applications to linear models, logistic and ordinal regression, and survival analysis*. Springer.

Hastie, T., & Tibshirani, R. (1986). Generalized additive models. *Statistical Science*, *1*(3), 297–310. https://doi.org/10.1214/ss/1177013604

Hastie, T., & Tibshirani, R. (1999). *Generalized additive models*. Chapman & Hall/CRC.

Haughton, J., & Khandker, S. R. (2009). *Handbook on poverty and inequality*. The World Bank. https://doi.org/10.1596/978-0-8213-7613-3

Hausmann, R., Pietrobelli, C., & Santos, M. A. (2020). Place-specific determinants of income gaps: New sub-national evidence from chiapas, mexico. *CID Working Paper Series*. https://dash.harvard.edu/handle/1/37366379

Heidinger, L. (2021). Intimate partner violence: Experiences of first nations, métis and inuit women in canada, 2018. *Juristat*, *85*(002-X).

Heise, L. (2011). *What works to prevent partner violence? an evidence overview*.

Hofner, B., Kneib, T., & Hothorn, T. (2016). A unified framework of constrained regression. *Statistics and Computing*, *26*(1-2), 1–14. https://doi.org/10.1007/s11222-014-9520-y

Hofner, B., Mayr, A., Robinzonov, N., & Schmid, M. (2014). Model-based boosting in r: A hands-on tutorial using the r package mboost. *Com-*

putational Statistics, *29*(1-2), 3–35. https://doi.org/10.1007/s00180-012-0382-5

Hothorn, T., Bühlmann, P., Kneib, T., Schmid, M., & Hofner, B. (2020). Mboost: Model-based boosting: R package version 2.9-4.

IFAD. (2016). Reducing rural women's domestic workload through labour-saving technologies and practices | fao.

ILO (Ed.). (2008). *Promotion of rural employment for poverty reduction*. https://www.ilo.org/ilc/ReportsavailableinGerman/WCMS_091721/lang--en/index.htm

ILO. (2020). The covid-19 response: Getting gender equality right for a better future for women at work: Policy brief. https://www.ilo.org/wcmsp5/groups/public/---dgreports/---gender/documents/publication/wcms_744685.pdf

INEGI. (2015a). Censo nacional de gobiernos municipales y delegacionales. https://www.inegi.org.mx/programas/cngmd/2015/

INEGI. (2015b). Encuesta intercensal. https://www.inegi.org.mx/programas/intercensal/2015/

INEGI. (2015c). Encuesta nacional de calidad e impacto gubernamental (encig). https://www.inegi.org.mx/programas/encig/2015/#Documentacion

INEGI. (2016a). Encuesta nacional de ingresos y gastos de los hogares (enigh): Nueva serie. https://www.inegi.org.mx/programas/enigh/nc/2016/

INEGI. (2016b). Encuesta nacional de victimización y percepción sobre seguridad pública (envipe). https://www.inegi.org.mx/

INEGI. (2016c). Encuesta nacional sobre la dinámica de las relaciones en los hogares (endireh). https://www.inegi.org.mx/

INEGI. (2019). *Mujeres y hombres en méxico*.

INEGI. (2020). Encuesta nacional de ingresos y gastos de los hogares (enigh). https://www.inegi.org.mx/programas/enigh/nc/2020/

INEGI. (2021). Encuesta nacional de ocupación y empleo. https://www.inegi.org.mx/programas/enoe/15ymas/

INEGI. (2022). Mortalidad. https://www.inegi.org.mx/programas/mortalidad/

Jaen Cortés, C. I., Rivera Aragón, S., Amorin de Castro, Elga Filipa, & Rivera Rivera, L. (2015). Violencia de pareja en mujeres: Prevalencia y factores asociados. *Acta de Investigación Psicológica*, *5*(3), 2224–2239. https://doi.org/10.1016/S2007-4719(16)30012-6

Johnstone, I. M., & Titterington, D. M. (2009). Statistical challenges of high-dimensional data. *Philosophical transactions. Series A, Mathematical,*

physical, and engineering sciences, *367*(1906), 4237–4253. https://doi.org/10.1098/rsta.2009.0159

Kan, Â. M. Y., & Laurie, H. (2016). Gender, ethnicity and household labour in married and cohabiting couples in the uk. *ISER Working Paper Series*, (2016-01). https://ideas.repec.org/p/ese/iserwp/2016-01.html

Khan, M. H. (2001). Rural poverty in developing countries implications for public policy.

Killewald, A., & Gough, M. (2010). Money isn't everything: Wives' earnings and housework time. *Social Science Research*, *39*(6), 987–1003. https://doi.org/10.1016/j.ssresearch.2010.08.005

Kim, S. M. (2014). The impacts of gender differences in social capital on microenterprise business start-up. *Affilia*, *29*(4), 404–417. https://doi.org/10.1177/0886109913519789

Klärner, A., & Knabe, A. (2019). Social networks and coping with poverty in rural areas. *Sociologia Ruralis*. https://doi.org/10.1111/soru.12250

Kneib, T., Hothorn, T., & Tutz, G. (2009). Variable selection and model choice in geoadditive regression models. *Biometrics*, *65*(2), 626–634. https://doi.org/10.1111/j.1541-0420.2008.01112.x

Kolpashnikova, K., & Koike, E. T. (2021). Educational attainment and housework participation among japanese, taiwanese, and american women across adult life transitions. *Asian Population Studies*, *17*(3), 266–284. https://doi.org/10.1080/17441730.2021.1920147

Krug, E. G., Dahlberg, L. L., Mercy, J. A., Zwi, A. B., & Lozano, R. (2002). World report on violence and health. https://doi.org/10.15496/publikation-8582

La Fuente, A. d. (2010). Remittances and vulnerability to poverty in rural mexico. *World Development*, *38*(6), 828–839. https://doi.org/10.1016/j.worlddev.2010.02.002

Lakner, C., Mahler, D. G., Negre, M., & Prydz, E. B. (2020). *How much does reducing inequality matter for global poverty?* World Bank, Washington, DC. https://doi.org/10.1596/33902

Lauritsen, J., & Schaum, R. (2004). The social ecology of violence against women. *Criminology*, *42*(2), 323–357. https://doi.org/10.1111/j.1745-9125.2004.tb00522.x

López Rosales, F., Moral de la Rubia, José, Diaz Loving, R., & Cienfuegos Martínez, Y. I. (2013). Violencia en la pareja. un análisis desde una perspectiva ecológica. *CIENCIA ergo-sum*, *20*(1), 6–16.

Lopez-Feldman, A., Mora, J., & Taylor, J. E. (2007). Does natural resource extraction mitigate poverty and inequality? evidence from rural mexico. *Environment and Development Economics*, *12*(2), 251–269. https://doi.org/10.1017/S1355770X06003494

Mavisakalyan, A., & Rammohan, A. (2021). Female autonomy in household decision-making and intimate partner violence: Evidence from pakistan. *Review of Economics of the Household*, *19*(1), 255–280. https://doi.org/10.1007/s11150-020-09525-8

Mckinley, T., & Alarcón, D. (1995). The prevalence of rural poverty in mexico. *World Development*, *23*(9), 1575–1585. https://doi.org/10.1016/0305-750X(95)00066-L

McManus, P. A., & DiPrete, T. A. (2001). Losers and winners: The financial consequences of separation and divorce for men. *American Sociological Review*, *66*(2), 246. https://doi.org/10.2307/2657417

Médor, D. (2013). Divorcio, discriminación y autopercepción en un grupo de mujeres en guadalajara, jalisco. *Papeles de población*, *19*(78), 41–64.

Meinshausen, N., & Bühlmann, P. (2010). Stability selection. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *72*(4), 417–473. https://doi.org/10.1111/j.1467-9868.2010.00740.x

Mojarro-Iñiguez, M., Valdez-Santiago, R., Pérez-Núñez, R., & Salinas-Rodríguez, A. (2014). No more! women reporting intimate partner violence in mexico. *Journal of Family Violence*, *29*(5), 527–537. https://doi.org/10.1007/s10896-014-9610-9

Mora-Rivera, J., & García-Mora, F. (2018). Microfinanzas y pobreza rural en méxico: Un análisis con técnicas de propensity score matching. *Cuadernos de Desarrollo Rural*, *15*(82), 1–19. https://doi.org/10.11144/Javeriana.cdr15-82.mprm

National Center for Injury Prevention and Control. (2020). Risk and protective factors|intimate partner violence|violence prevention. Retrieved, from https://www.cdc.gov/violenceprevention/intimatepartnerviolence/riskprotectivefactors.html

Nguyen, T. H., van Ngo, T., Nguyen, V. D., Nguyen, H. D., Nguyen, H. T. T., Gammeltoft, T., Wolf Meyrowitsch, D., & Rasch, V. (2018). Intimate partner violence during pregnancy in vietnam: Prevalence, risk factors and the role of social support. *Global health action*, *11*(sup3), 1638052. https://doi.org/10.1080/16549716.2019.1638052

Nitsche, N., & Grunow, D. (2016). Housework over the course of relationships: Gender ideology, resources, and the division of housework from

a growth curve perspective. *Advances in Life Course Research*, *29*, 80–94. https://doi.org/10.1016/j.alcr.2016.02.001

Oduro, A. D., Deere, C. D., & Catanzarite, Z. B. (2015). Women's wealth and intimate partner violence: Insights from ecuador and ghana. *Feminist Economics*, *21*(2), 1–29. https://doi.org/10.1080/13545701.2014.997774

Oficina de Violencia Doméstica. (2021). Informe estadístico anual: Año 2020.

Olesen, T. B., Jensen, K. E., Nygård, M., Tryggvadottir, L., Sparén, P., Hansen, B. T., Liaw, K.-L., & Kjaer, S. K. (2012). Young age at first intercourse and risk-taking behaviours–a study of nearly 65 000 women in four nordic countries. *European journal of public health*, *22*(2), 220–224. https://doi.org/10.1093/eurpub/ckr055

Ordaz, J. L. (2009). *México: Impacto de la educación en la pobreza rural.* CEPAL. https://repositorio.cepal.org/handle/11362/4883

Otega, O., & Muneer'deen, M. (2014). Good governance, rural development and poverty alleviation in nigeria: Issues and challenges. *International Journal of Sustainable Development & World Policy*, *3*(4), 100–114. https://ideas.repec.org/a/pkp/ijsdwp/v3y2014i4p100-114id2037.html

Peek-Asa, C., Saftlas, A. F., Wallis, A. B., Harland, K., & Dickey, P. (2017). Presence of children in the home and intimate partner violence among women seeking elective pregnancy termination. *PLOS ONE*, *12*(10). https://doi.org/10.1371/journal.pone.0186389

Plazaola-Castaño, J., Ruiz-Pérez, I., & Isabel Montero-Piñar, M. (2008). Apoyo social como factor protector frente a la violencia contra la mujer en la pareja. *Gaceta sanitaria*, *22*(6), 527–533. https://doi.org/10.1016/s0213-9111(08)75350-0

Priorities and strategies for education: A world bank review. (1995). World Bank.

Rapp, D., Zoch, B., Khan, M. M. H., Pollmann, T., & Krämer, A. (2012). Association between gap in spousal education and domestic violence in india and bangladesh. *BMC public health*, *12*(1), 467. https://doi.org/10.1186/1471-2458-12-467

Rashada, A. S., & Sharaf, M. F. (2016). Income inequality and intimate partner violence against women: Evidence from india.

Reichel, D. (2017). Determinants of intimate partner violence in europe: The role of socioeconomic status, inequality, and partner behavior. *Journal*

of interpersonal violence, *32*(12), 1853–1873. https://doi.org/10.1177/0886260517698951

Rivera-Rivera, L., Lazcano-Ponce, E., Salmerón-Castro, J., Salazar-Martínez, E., Castro, R., & Hernández-Avila, M. (2004). Prevalence and determinants of male partner violence against mexican women: A population-based study. *Salud publica de Mexico*, *46*(2), 113–122. https://doi.org/10.1590/s0036-36342004000200005

Rubiano Matulevich, E. C., & Viollaz, M. (2019). Gender differences in time use: Allocating time between the market and the household. https://openknowledge.worldbank.org/handle/10986/32274

Samtleben, C., & Müller, K.-U. (2022). Care and careers: Gender (in)equality in unpaid care, housework and employment. *Research in Social Stratification and Mobility*, *77*, 100659. https://doi.org/10.1016/j.rssm.2021.100659

Schmid, M., & Hothorn, T. (2008). Boosting additive models using component-wise p-splines. *Computational Statistics Data Analysis*, *53*(2), 298–311. https://doi.org/https://doi.org/10.1016/j.csda.2008.09.009

Shah, R. D., & Samworth, R. J. (2013). Variable selection with error control: Another look at stability selection. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *75*(1), 55–80. https://doi.org/10.1111/j.1467-9868.2011.01034.x

Skoufias, E., Lunde, T., & Patrinos, H. A. (2010). Social networks among indigenous peoples in mexico. *Latin American research review*, *45*(2), 49–67.

Sterling, S. (2018). Intimate partner violence in mexico: An analysis of the intersections between machismo culture, government policy,and violence against women. *International Studies Undergraduate Honors Theses*.

Stöckl, H., Hassan, A., Ranganathan, M., & M Hatcher, A. (2021). Economic empowerment and intimate partner violence: A secondary data analysis of the cross-sectional demographic health surveys in sub-saharan africa. *BMC Women's Health*, *21*(1), 241. https://doi.org/10.1186/s12905-021-01363-9

Stöckl, H., March, L., Pallitto, C., & Garcia-Moreno, C. (2014). Intimate partner violence among adolescents and young women: Prevalence and associated factors in nine countries: A cross-sectional study. *BMC public health*, *14*(1), 1–14. https://doi.org/10.1186/1471-2458-14-751

Stöckl, H., Watts, C., & Penhale, B. (2012). Intimate partner violence against older women in germany: Prevalence and associated factors. *Journal of interpersonal violence*, *27*(13), 2545–2564. https://doi.org/10.1177/0886260512436390

Sturge-Apple, M. L., Skibo, M. A., & Davies, P. T. (2012). Impact of parental conflict and emotional abuse on children and families. *Partner Abuse*, *3*(3), 379–400. https://doi.org/10.1891/1946-6560.3.3.379

Terrazas-Carrillo, E. C., & McWhirter, P. T. (2015). Employment status and intimate partner violence among mexican women. *Journal of interpersonal violence*, *30*(7), 1128–1152. https://doi.org/10.1177/0886260514539848

Torres Munguía, J. A., & Martínez-Zarzoso, I. (2020). What determines poverty in mexico? a quantile regression approach.

Torres Munguía, J. A., & Martínez-Zarzoso, I. (2021). Examining gender inequalities in factors associated with income poverty in mexican rural households. *PloS one*, *16*(11), e0259187. https://doi.org/10.1371/journal.pone.0259187

Torres Munguía, J. A., & Martínez-Zarzoso, I. (2022). Determinants of emotional intimate partner violence against women and girls with children in mexican households: An ecological framework [PMID: 35135364]. *Journal of Interpersonal Violence*, *0*(0), 08862605211072179. https://doi.org/10.1177/08862605211072179

UN (Ed.). (2016). *Women's economic empowerment in the changing world of work: Report of the secretary-general: Follow-up to the fourth world conference on women and to the twenty-third special session of the general assembly, entitled "women 2000: Gender equality, development and peace for the twenty-first century": Implementation of strategic objectives and action in critical areas of concern and further actions and initiatives* (E/CN.6/2017/1). https://www.un.org/ga/search/view_doc.asp?symbol=E/CN.6/2017/3

UN Women. (2018). *Turning promises into action: Gender equality in the 2030 agenda for sustainable development*. UN Women.

UN Women. (2019). *Progress of the world's women 2019-2020: Families in a changing world*. United Nations. https://doi.org/10.18356/696a9392-en

UNDP. (2019). Informe de desarrollo humano municipal 2010-2015. transformando méxico desde lo local | el pnud en méxico. https://www.mx.undp.org/

UNiTE Working Group. (2019). A resource book on intimate partner violence for united nations staff in asia.

Valdez-Santiago, R., Híjar, M., Rojas Martínez, R., Avila Burgos, L., & La Arenas Monreal, M. d. L. (2013). Prevalence and severity of intimate partner violence in women living in eight indigenous regions of mexico. *Social science & medicine (1982)*, *82*, 51–57. https://doi.org/10.1016/j.socscimed.2013.01.016

Verner, D. (2005). *Poverty in rural and semi-urban mexico during 1992-2002*. The World Bank. https://doi.org/10.1596/1813-9450-3576

Villarreal, A. (2007). Women's employment status, coercive control, and intimate partner violence in mexico. *Journal of Marriage and Family*, *69*(2), 418–434. https://doi.org/10.1111/j.1741-3737.2007.00374.x

Voith, L. A., Azen, R., & Qin, W. (2021). Social capital effects on the relation between neighborhood characteristics and intimate partner violence victimization among women. *Journal of urban health : bulletin of the New York Academy of Medicine*, *98*(1), 91–100. https://doi.org/10.1007/s11524-020-00475-1

Walton-Moss, B. J., Manganello, J., Frye, V., & Campbell, J. C. (2005). Risk factors for intimate partner violence and associated injury among urban women. *Journal of Community Health*, *30*(5), 377–389. https://doi.org/10.1007/s10900-005-5518-x

WHO. (2012). *Understanding and addressing violence against women: Intimate partner violence*.

Wilson, N. (2019). Socio–economic status, demographic characteristics and intimate partner violence. *Journal of International Development*, *31*(7), 632–657. https://doi.org/10.1002/jid.3430

Wood, S. N. (2017). *Generalized additive models: An introduction with r* (Second edition). CRC Press/Taylor & Francis Group.

World Bank. (2005). *Income generation and social protection for the poor: Volume 4. a study of rural poverty in mexico*. Washington, DC.

Wright, E. M. (2015). The relationship between social support and intimate partner violence in neighborhood context. *Crime & Delinquency*, *61*(10), 1333–1359. https://doi.org/10.1177/0011128712466890

Yilmaz, O. (2018). Female autonomy, social norms and intimate partner violence against women in turkey. *The Journal of Development Studies*, *54*(8), 1321–1337. https://doi.org/10.1080/00220388.2017.1414185

Zhu, N., & Luo, X. (2010). The impact of migration on rural poverty and inequality : A case study in china. *01695150*. https://openknowledge.worldbank.org/handle/10986/5043