MASTERS

APPLIED ECONOMETRICS AND FORECASTING

MASTER'S FINAL WORK

2023/2024

DISSERTATION

THE EFFECT OF TEENAGE PREGNANCY ON ECUADORIAN WOMEN'S YEARS OF SCHOOLING

EDUARDO SEBASTIAN HARO PAREDES

SUPERVISOR:

Professor ADRIANA CORNEA-MADEIRA, PH.D.

JUNE – 2024

Abstract

Teenage pregnancy is a complex issue with significant consequences, particularly in Latin America, where Ecuador experiences one of the highest rates among South American countries. This study aims to determine the causal effect of teenage pregnancy on the educational attainment of adolescent girls in Ecuador. Using data from the "Encuesta Nacional de Salud y Nutricion" ENSANUT 2018 and incorporating the Double/Debiased Machine Learning methodology alongside the Instrumental Variables framework, the study finds that teenage mothers have, on average, 4 to 6 less years of schooling compared to non-teenage mothers. This implies that women with teenage pregnancies have a lower educational level than women with non-teenage pregnancies and are therefore more susceptible to social problems such as higher unemployment, lower income, lower life quality, among others. The results offer valuable insights for policymakers, highlighting the need for targeted interventions like flexible education programs, support services, financial assistance, and more to address the educational deficits faced by teenage mothers, thereby improving their quality of life and prospects.

Keywords: education, teenage pregnancy, menarche, instrumental variables, double machine learning

To Victoria and Polivio, without them I would be nothing.

Acknowledgments

I would like to thank my friends from the master, who I consider my family here in Portugal. To my friends in Ecuador who have helped me not to feel so far away from home. To Andreia and Alex, who I consider my Portuguese sister and my Ecuadorian brother respectively. To Margarida, for her warm and kind soul, that has always been a source of comfort and peace in tough times. To my family for their constant help and unconditional support throughout this adventure. To Toya, Gabriela, and Juan Carlos who always look out for my welfare and never hesitate to help me when I need it. To Pamela and Rick, who I don't see them very often, but they have always been there for me, providing me with their affection, guidance, and assistance in everything I propose to myself. Finally, to my parents, Victoria and Polivio, for giving me all their love, affection, and unconditional support throughout my life. Without all of you, I would not have made it this far. Thank you all.

Table of Tables

Table of Figures

List of Abbreviations

ATE – Average Treatment Effects

LATE – Local Average Treatment Effects

IV – Instrumental Variables

FE – Fixed Effects

OLS – Ordinary Least Squares

TOT – Treatment On the Treated

TSLS – Two Stage Least Squares

LIML – Limited Information Maximum Likelihood

ML – Machine Learning

DML – Double/Debiased Machine Learning

RF – Random Forests

BAG – Bagging

SVR – Support Vector Regression

RBF – Radial Basis Function

MLP – Multi-Layer Perceptron Regressor

NN – Neural Network

ReLu – Rectified Linear Unit

ENSANUT – Encuesta Nacional de Salud Nutricion

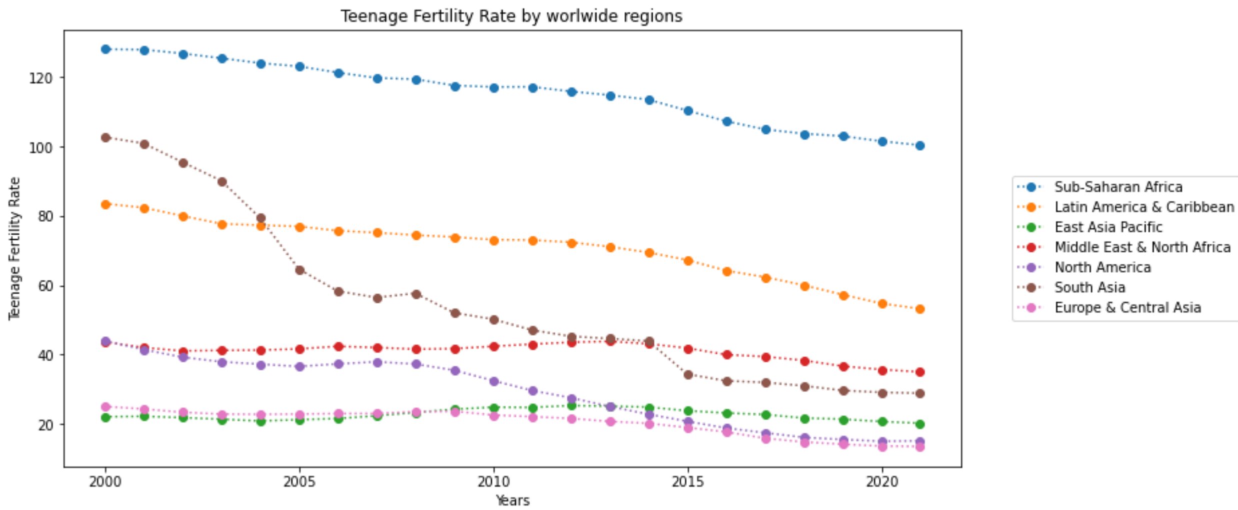
Table of Contents

1. Introduction

Teenage childbearing is a problem with diverse consequences for women in different aspects. For instance, according to Hofferth (1987), early fertility seems to be related to a variety of unfavorable economic consequences, including decreased earnings and family revenue, increased poverty rates, and an increased likelihood of reliance on welfare. Furthermore, Geronimus & Korenman (1992) stated that teenage childbearing is a cause of persistent poverty and poverty transmitted intergenerationally. Moreover, early research (Card & Wise, 1978; Hayes, 1987; Mott & Marsiglio, 1985) suggests that bearing children when underage significantly decreases educational attainment. This indicates that women are forced to avoid investing in their own education and training due to the substantial costs of motherhood (Becker, 1981; Manlove, 1998). The World Health Organization (n.d.) defines adolescence as the phase in life between childhood and adulthood, from ages 10 to 19 and this term is used interchangeably with the term teenage and all its derivatives in this study.

In Latin America and the Caribbean, teenage pregnancy has become a matter of concern in public policy, since this region has the second highest teenage pregnancy rate in the world in 2021, only surpassed by Sub-Saharan Africa (World Bank, 2024). Figure 1 shows that in 2000, Latin America and the Caribbean had the third highest adolescent fertility rate (births per 1000 woman from 15 to 19 years old) among the 7 world regions. Although rates are declining worldwide, Latin America's extraordinarily slow decline compared to other regions explains the region's gradual rise to the top of the adolescent fertility charts.

Figure 1. Teenage Fertility Rates by worldwide region



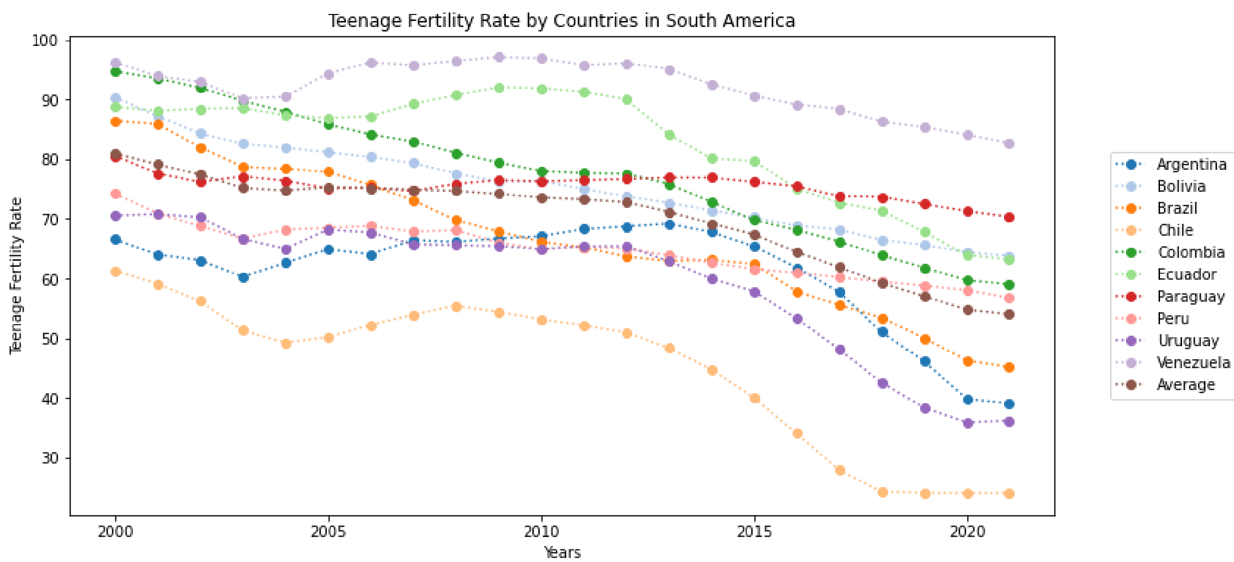
This high level of adolescent pregnancy translates into a variety of social problems for Latin American women. A study conducted by the United Nations Population Fund (UNFPA, 2020) in Argentina, Colombia, Ecuador, Guatemala, Mexico, and Paraguay shows that women who became early mothers were 3 times less likely to get a university degree as adults. On average, 6.4% of women who became mothers during adolescence attained the same degree of higher education as did 18.6% of women who were mothers in the first ten years of their adult lives (UNFPA, 2020).

In the job market, women who become moms later in life make 24% more money than teenage mothers. On average, the yearly labor income for teenage mothers is USD 3,068 whereas it is USD 4,015 for adult mothers. Also, a higher proportion of women who became early mothers are exclusively dedicated to unpaid domestic work since teenage mothers showed a 46.8% of job inactivity in contrast to the 41.9% of inactivity in adult mothers. Furthermore, teenage mothers show higher unemployment rates than those adult mothers, having a 6.8% vs 5.1% rates respectively (UNFPA, 2020).

For the government, the combined annual tax revenue lost by the six researched countries' governments comes to USD 746 million in 2018, or USD 110 for each adolescent pregnant woman where 56% is estimated to account for Valued Added Tax (VAT) and 44% to Income Tax (IT). Also, government expenditure for pregnancy delivery and puerperium care for pregnant teenagers ranges from USD 6.7 million to USD 305 million depending on the country. Additionally, the amount of welfare spending that could be saved by using preventative measures ranges from USD 4.8 million to USD 211 million (UNFPA, 2020).

Although the teenage pregnancy rate has gradually decreased in Latin America and the Caribbean, in South America, the trends shifts are different per country. Figure 2[1] shows the evolution of fertility rates in South America where countries like Chile, Uruguay, Brazil, and Argentina have reduced significatively their fertility rates to the point of being below the region's average with a 60%, 48%, 47% and 41% respectively. On the other side, countries like Paraguay, Venezuela, Ecuador, and Bolivia have not made significant progress to reduce it with 12%, 14%, 28% and 29% respectively. It is also noteworthy that decrease of the average fertility rate of South America is similar to the decrease Latin-American's rate having a 33% vs 36% respectively (World Bank, 2024).

Figure 2. Teenage Fertility Rates in South America



For Ecuador, it is quite evident that teenage pregnancy is still a significant problem that has not been properly addressed by any government and due to the lack of research of its consequences in women, there is no way to measure its impact in different aspects of women's life (e.g., years of schooling). For this, the main objective of this study is to assess the relationship between teenage pregnancy and the years of schooling for

[1] Guyana, Suriname, and French Guyana were not considered.

Ecuadorian women using econometric methodologies, and thus contribute with statistical evidence that can be used for initiating the debate at a national level that encourages the government to invest resources for the design and implementation of effective public policies to reduce this multidimensional problem.

The remainder of this study is organized as follows. Section 2 provides the literature review that studied the effects of teenage pregnancy on educational attainment. Section 3 summarizes the main methodology that was implemented. Section 4 describes the source of the data used in the research alongside with some stylized facts and descriptive statistics that stem from it. Section 5 presents the main results, and finally Section 6 provides the main conclusions, contributions, limitations, and future research of this study.

2. Literature Review

There is abundant literature that has examined the relationship between fertility timing and education attainment using diverse methodologies and assumptions. This literature can be classified in three broad groups. The first-generation of studies were multivariate analysis that considered fertility as an exogenous determinant of education attainment. Moreover, the second-generation used the Instrumental Variables (IV) methodology to address the endogeneity of fertility. Finally, the third-generation of studies used various techniques like Fixed Effects (FE), quasi-natural experiments, matching techniques, and survival analysis to ameliorate statistical problems and timing effects more closely (Ribar, 1994).

For the first-generation studies, regression-approach methodologies were used to estimate the Average Treatment Effect (ATE) where the schooling years were regressed on a teen pregnancy indicator variable (D_i) and a vector of covariates (X_i) that were thought to possible confound the causal effect. Early results from this research suggest considerable reduction in educational attainment in teen mothers, ranging from 2 to 4 less years of schooling (Card & Wise, 1978; Moore & Waite, 1977; Mott & Marsiglio, 1985; Waite & Moore, 1978). However, this approach has been widely abandoned in favor of other methods that relax functional form assumption and consider endogeneity in the teen pregnancy indicator variable.

To account for this endogeneity, the second-generation of studies employ the IV methodology. Essentially, IV estimates seek to determine a causal effect through the exogenous variation in teenage pregnancies (Ribar, 1999). The exogenous variation comes from an instrument that must comply with two conditions: relevance and exogeneity (Wooldridge, 2002).

Several studies have tried to estimate the causal effect of teenage pregnancy through IV methodology using miscarriages as the most common instrument (Ashcraft et al., 2013; Ashcraft & Lang, 2006; Fletcher & Wolfe, 2009; Rindfuss et al., 1980). With exceptions of decreased fertility, miscarriages are thought to be randomly distributed in relation to socioeconomic traits on average. However, this approach presents various limitations. First, this method restricts generalizability to teenage pregnancies, which is already a restricted population, moreover, the IV estimate fails to account for young women who would terminate their pregnancy or woman who would like to seek adoption in the absence of a miscarriage, or who would seek another child after a loss. Thus, estimates are likely to be Treatment On the Treated (TOT) effects (Diaz & Fiel, 2016).

Although miscarriages are a common instrument used in several studies, other authors have proposed the age of menarche as another possible instrument. Stock & Watson (2018) suggested two methods for the selection of valid instruments. First, is the use of economic theory as a selection mechanism and second is the search for some exogenous source of variation in a instrument arising from what is, in effect, a random phenomenon that induces changes in the endogenous regressor. Following this, Presser (1978) suggests that the age of menarche can be seen as a indicator of the timing of fecundity and socio-sexual behavior. Thus, it is plausible that a younger age of menarche influences teenage childbearing by lengthening the exposure to fertility.

Ribar (1994) developed and estimated a simultaneous discrete choice model, a bivariate probit, of adolescent fertility measured as childbearing before age of 20 and high school completion using the age of menarche as a instrument. His results suggest that teen pregnancy reduces the probability of completing high school by 23.4% assuming exogeneity of early fertility, however, after accounting for endogeneity, his results are not statistically significant. Because of this, he tried different specifications for teenage pregnancy yielding similar results, statisitically significant and negative coefficients assuming exogeneity and non-statistical significant coefficients assuming endogeneity. Similarly, Chevalier & Viitanen (2002) estimated the effect of teenage motherhood as the the decision to invest in post-compulsory education and their qualifications level at age 33 using IV and matching techniques. Their results suggests that teenage pregnancy significantly impairs education and lowers the likelihood of completing post-compulsory education by 24%, matching estimates, however, reduce this disparity to a still significant 12%–17%. Their results suggest that teenage moms have a persistent educational disadvantage, making them less competent than other women by the time they are 33 and higlight the challenges of managing a small child's needs while attending school, as well as the dearth of opportunities for adult education.

Keplinger et al. (1999) studied the negative effects of early childbearing on typical human capital investment activities within adolescence and early adulthood (i.e., high school completion, college entrance, work experience) using the age of menarche as an instrument. The results showed that on average white women who did not avoid early childbearing completed 2.4 fewer years of schooling, had 0.6 fewer years of early work experience, and had a median hourly wage $2.09 lower than those who did not become mothers before age 19. The results for black women are 1.6 fewer years of education, 0.3 fewer years of early work experience, and lower earnings by $1.19 per hour.

For the third-generation studies, Ribar (1999) studied the effects of teen pregnancy on years of schooling using FE estimators to account for unobserved characteristics shared among sisters. The reasoning behind this approach is to account for unknown family-specific factors that influence early fertility and educational achievement, that could potentially bias the estimated treatment effect. The treatment variability required to identify effects stems from families where at least one sibling experienced teenage pregnancy and at least one did not. FE models provide unbiased treatment effects as long as this variation is conditionally uncorrelated with unmeasured sibling differences that impact achievement.

However, this metholodogy has some limitations. First, reduced generalizability by taking advantage of between sibling variance limits the sample to households with several daughters, moreover, identification relies on households where the treatment varies. Therefore, a disproportionate share of these estimates comes from young women from bigger households who have greater rates of teen pregnancy than the general population. FE models can also be seen as OLS models controlling for family identifiers; for this, FE methodologies estimate variance-weighted average treatment effects, giving families with higher treatment variation more weight. As a result, treatment effects in sibling FE models are probably close to the TOT (Diaz & Fiel, 2016).

FE estimates are generally smaller than OLS estimates in the literature (Geronimus & Korenman, 1992). Estimates of FE methodology falls between 0 and a 1 year reduction in schooling for teenage mothers (Hoffman et al.,1993; Holmund, 2005).

Another approach is the use of Propensity Score Matching technique to examine treatment effects. Briefly, two steps address selection into treatment. First, a logit model is used to estimate the probability that a woman will become pregnant as a teenager. A set of pretreatment covariates is considered to make the treatment conditionally independent of outcomes. Second, predicted propensity scores are used to establish matched pairs of treatment and control cases. Then, treatment effects are derived by essentially averaging the sample's within-pair differences (Diaz & Fiel, 2016).

Several studies have used Propentsity Score matching technique to estimate the effects of teen childbearing. For instance, research from (Chevalier & Viitanen, 2002; Lee, 2010) estimated ATE where results suggest that teen mothers are almost 40% less likely to complete college. In contrast, others (Levine & Painter, 2003; Sanders & Zhang, 2007) focused on TOT effects and estimated smaller effect's on women's subsequent schooling.

A different approach was used by Olsen & Farkas (1989) who estimated survival models of the impact of childbearing on the time to high school leave using experimental data. Their results suggest thart the age at first birth had no apparent effect on education, which took into consideration the potential endogeneity of fertility timing. Similarly, Upchurch & McCarthy (1990) also used survival models and found that fertility had no significant effect on women's decisions to school dropout but it has significant negative effects on drop-outs' decisions to return to school.

Finally, Grogger & Bronars (1993) exploited the probability of having twins, contingent on getting pregnant, to set up a natural experiment. They found that there are long-term effects on the likelihood of future marriage and family earnings, but only for black people, and significant effects on the short-term labor force involvement of all teenage mothers.

This paper will use the second-generation approach as the main methodological framework, more specifically, the use of IV and the age of menarche as the only instrument. However, unlike the previously mentioned, this paper will have different settings. First, this study will be a cross-sectional study instead of the panel data study mentioned before. Second, it will also incorporate IV techniques alongside Machine Learning and finally, due to the structure of the data, the analysis will be done under different covariates settings.

3. Methodology

3.1 IV Framework

As previously stated, the main objective of this research is to determine the causal relationship between teenage pregnancy and the years of schooling of Ecuadorian women. However, it is believed that teenage pregnancy presents endogeneity problems that do not allow for unbiased estimates of its effect on years of schooling. Therefore, to solve this problem, IV methodology was employed to counteract the effects of endogeneity.

For an instrument to be valid, it must meet two conditions: relevance and exogeneity. These characteristics allow the estimation of the model to capture all movements of the variable of interest free of endogeneity. These two conditions can be represented by the formulae:

1. Relevance: $Cov(Z_i, T_i) \neq 0$
2. Exogeneity: $Cov(Z_i, u_i) = 0$

where Z_i is the instrument, T_i is the endogenous variable and u_i is the error term associated to a regression equation with Y_i as the dependent variable alongside the observed confounders. If the instrument is relevant, then the variability of the instrument is related to the variability of an endogenous variable and therefore, related to the variability of the dependent variable. If the instrument is exogenous, then variability of the endogenous variable captured by the instrument is exogenous. Thus, if the instrument meets both conditions, it captures the variability of the endogenous variable that is exogenous (Stock & Watson, 2018).

If an instrument satisfies both conditions, the causal effect can be estimated using the Two Stage Least Squares (TSLS). As suggested by the name, this estimator is calculated in two stages. The first one, decomposes the variation of the endogenous variable in a component that is correlated with u_i and another that is uncorrelated with u_i. The second one uses the uncorrelated variation to estimate the unbiased effect. The TSLS methodology can be expressed by the following system of equations:

$$Y_i = \theta_0 + \theta_1 T_i + \theta_2 X_i + U_i$$

$$T_i = \beta_0 + \beta_1 Z_i + \beta_2 X_i + V_i$$

where for our case, Y_i is the schooling years, T_i is the indicator variable for a teenage pregnancy, Z_i is the age of menarche as the instrument, X_i is a set of covariates, U_i is the error term of the second stage and V_i is the error term of the first stage.

The two-stage model can be easily estimated by first, regressing the IV and covariates on the endogenous variable and generate prediction values.

$$\hat{T}_i = \hat{\beta}_0 + \hat{\beta}_1 Z_i + \hat{\beta}_2 X_i$$

The next step is to use predicted values of $\hat{T}_i$ and regress them alongside the covariates to estimate an accurate TSLS estimator $\hat{\theta}_{TSLS}$.

$$y_i = \hat{\theta}_0 + \hat{\theta}_{TSLS}\, \hat{T}_i + \hat{\theta}_2 X_i$$

The formula for the TSLS estimator is:

$$\hat{\theta}_{TSLS} = (Z'X)^{-1} Z'Y$$

If the instrument is valid, then the TSLS estimator is consistent and in large samples, it is normally distributed. Because of this, it is possible to state that:

$$\hat{\theta}_{TSLS} \sim N(\theta_1, \sigma^2{}_{TSLS})$$

where $\sigma^2{}_{TSLS}$ = $\sigma^2 (Z'X)^{-1} Z'Z (X'Z)^{-1}$.

To test the validity of the model, the endogenous test and weak instrument test were performed, however, the overidentification test was not performed since the model is exactly identified because there is one endogenous variable and one instrument. For the relevance condition, Staiger & Stock (1997) proposed general and empirical rule of thumb, that the F-statistic of the first stage of the TSLS estimator should be greater than 10 for the instrument to be considered not weak. Stock & Yogo (2005) stated that this rule of thumb is motivated based on the rule of relative bias of the TSLS when the relative bias is approximately 10% with a 5% significance level with only one endogenous regressor. They generalized this idea by formally proposing a characterization for weak instruments for a general number of endogenous regressors and providing a test to formally determine if an instrument is weak or not. However, their test relies on homoscedastic errors assumption, which is highly non-plausible.

Because of this, Montiel & Pflueger (2013) developed a test for weak instruments in linear IV regressions that is robust to heteroscedasticity, autocorrelation, and clustering. This test is an extension of the test proposed by Stock & Yogo (2005) that tests the null hypothesis instruments are either weak or the estimator's (Nagar, 1959) bias is relatively large to benchmark for both TSLS and LIML with one endogenous regressor. This framework allows to use the Eicker-Huber-White heteroscedasticity robust estimates (Newey & West, 1987), heteroskedasticity and autocorrelation consistent estimates, and clustered variance estimates. Finally, for the endogeneity tests, the Durbin and Wu-Hausman test were used.

Although Two Stage Least Squares is a ubiquitous tool for causal estimation using IV, it also has its own downfalls. Following Imbens & Wooldridge (2009), any methodology that tries to estimate treatment effects using regressions heavily relies on the specification of a correct functional form of the model for the estimators to be unbiased. This means that linearity assumption and the correct inclusion a set of covariates alongside with their powers is key for identifying a correct functional form, thereby, an unbiased estimator (Chan & Mátyás, 2022). For cases where a proper functional form cannot be properly specified due to the previous reasons and it is desired to estimate causal effects, a new framework called Double Machine Learning helps to address this problem.

According to Moler-Zapata et al. (2023), TSLS estimates LATE under heterogeneity given a binary instrument. However, when the instrument is continuous, such as the case of the age of menarche, this method estimates the weighted average of LATEs.

3.2 Double/Debiased Machine Learning

3.2.1 Overview

Double Machine Learning (DML) is a methodology that estimates treatment effects when all potential confounders/controls are observed (unconfoundedness/ignorability assumption) but are either too many for classical statistical approaches to be applicable or their effect on the treatment and outcome cannot be modeled by parametric functions. Chernozhukov et al. (2018) proposed a methodology that allows to estimate causal or treatment effects parameters with the help of nuisance parameters that are estimated using different Machine Learning (ML) methods like random forests, lasso or post-lasso, neural nets, boosted regression trees, and various hybrids and ensembles of these methods. These ML methods can handle many covariates and model the nonlinear relationships between the covariates and the dependent variable.

In brief, Chernozhukov et al. (2018) provide a lead example of this methodology using a Partially Linear Regression (PLR) model that can be adapted to the problem in analysis with:

$$Y_i = \theta_0 T_i + \ g_0(X) + U_i \ , E[U_i \mid X, T_i] = 0$$

$$T_i = m_0(X) + V \ , E[V_i \mid X] = 0$$

where Y_i is the schooling years, T_i is indicator variable for teenage pregnancy, X_i is vector of p covariates $X = (X_{i,1}, X_{i,2}, X_{i,3}, \dots, X_{i,p})$ and U_i and V_i are disturbances. The main equation is the first one where θ_0 is the regression coefficient of main interest where if the treatment variable conditional on the controls is exogeneous, this parameter has a causal interpretation. Confounding is monitored by the second equation, which reflects the treatment variable's reliance on controls. Although this equation is not particularly interesting, it is crucial for identifying and eliminating regularization bias. The endogenous variable T_i is impacted by the confounding factors X through the function $m_0(X)$, while the outcome variable is impacted by the function $g_0(X)$.

A naive approach for the estimation of θ_0 would be to directly estimate the first equation by first fitting a ML method to estimate $\widehat{g_0(X)}$ and then estimate $\widehat{\theta_0}$ with OLS.

Although this framework seems straightforward, it has its own drawbacks. As mentioned before, DML can handle cases where a vast number of covariates are available through regularization techniques such as Lasso, Ridge, Elastic Net and many more. These estimators' regularization prevents the estimator's variance from exploding, but it also inevitably introduces significant biases in the estimator $\widehat{g_0(X)}$ of $g_0(X)$.

To overcome the regularization bias, a procedure called "Orthogonalization" is used. This procedure consists in partialling out the effect of X from T_i to obtain an orthogonalized regressor $V_i = T_i - m_0(X)$. Specifically, the goal is to estimate $\widehat{V_i} = D - \widehat{m_0(X)}$, where $\widehat{m_0(X)}$ is a ML estimator of $m_0(X)$. Because of the auxiliary prediction condition, the methodology is called "Double Machine Learning" or "Debiased Machine Learning".

Although orthogonalization helps to deal with regularization bias, it also induces another bias that arises from overfitting. But, unlike the regularization bias, this overfitting bias can be properly addressed with the cross-fitting approach. As an illustrative example, cross-fitting first starts by splitting the sample in two parts, the main and auxiliary sample, where the auxiliary sample is used to estimate the ML models and the first one is used to estimate $\theta_{0,1}$. Similarly, another estimate of θ_0 can be obtained by estimating the ML models in the main sample and estimating $\theta_{0,2}$ in the auxiliary sample. A final estimator of θ_0 is obtained by averaging $\theta_{0,1}$ and $\theta_{0,2}$. This new estimator is more robust to overfitting and the cross-fitting procedure can be extended to incorporate K sample for an even more robust estimator.

Though the previous framework correctly incorporates ML with a causal inference framework, it heavily relies on the unconfoundedness assumption. This assumption although plausible, is based on the fact all covariates/confounders are observed, which is highly unlikely in social studies. Fortunately, Chernozhukov et al. (2018) also provide an extension on this framework that incorporates IV. This extension also inherits the problems and solution from the previous framework.

The new model considers the following structure:

$$Y_i = \theta_0 T_i + g_0(X) + U_i \, , E[U_i \mid X, T_i] = 0$$

$$Z_i = m_0(X) + V_i \,, E[V_i \mid X] = 0$$

where for this research, Y_i is the years of schooling, T_i is an indicator variable that states the condition of a mother that had a teen pregnancy or not, Z_i is the IV that represents the age of menarche of the mother and X is a set of covariates listed in Section 4.2 and U_i and V_i are disturbances. To make inference on θ_0, this approach uses the Robinson-style "partialling-out" score function:

$$\psi(W; \theta, \eta) := \Big(Y_i - l(X) - \theta\big(T_i - r(X)\big)\Big)\big(Z_i - m(X)\big) \,, \eta = (l, m, r)$$

where $W = (Y_i, T_i, X, Z_i)$ and l, m and r are P-square-integrable functions that map the support of X to R. This score function satisfies the moment condition $E_P\, \psi(W; \theta_0, \eta_0) = 0$ and the orthogonality condition $\partial_\eta E_P \psi(W; \theta_0, \eta_0)[\eta - \eta_0] = 0$ for $\eta_0 = (l_0, m_0, r_0)$ for $l_0 = E_P[Y_i \mid X]$, $r_0 = E_P[T_i \mid X]$ and $m_0 = E_P[Z_i \mid X]$ and all functions in η_0 can be estimated using ML methods.

The first step to apply the Partially Linear IV model is to determine the ML models that will be used to estimate the causal parameter. For this study, this ML models will be Random Forests (RF), Bagging Trees (BAG), Support Vector Machines (SVM) and Multilayer Perceptron (MLP) alongside with the standard Linear Regression framework as benchmark, which turns out to be the TSLS, and details of their implementation are discussed later from Section 3.2.3 to 3.2.6. Moreover, like any ML model, these methods are prone to overfitting/underfitting if they are not trained properly. To address this, a hyperparameter tuning procedure was employed. Details of the implementation of the ML models and the hyperparameter tuning algorithms are discussed in Section 3.2.2. For l_0 and m_0, regression ML methods were implemented whereas for r_0, classification ML methods were used except for the OLS where all three functions are regressions methods.

3.2.2 Hyperparameter Tunning

For all ML models, the hyperparameter algorithm starts with the definition of the parameter grid, which is a structured way of defining a set of hyperparameters and their corresponding values for a ML model. Hyperparameter tuning is performed using the “GridSearchCV” from the Python library “Scikit-learn” (Pedregosa et al., 2011) .The algorithm can be summarized as the following:

- **Parameter Grid Generation:** It generates all possible combinations of hyperparameters specified in the parameter grid.
- **Cross-Validation:** It splits the training data into multiple folds (typically using k-fold cross-validation). For each combination of hyperparameters:
    - It trains the model on a subset of the data (training set).
    - It evaluates the model on the remaining subset (validation set).
- **Model Fitting and Evaluation:** For each combination of hyperparameters, it fits the model to the training data and evaluates its performance on the validation data using a specified scoring metric.
- **Best Model Selection:** It selects the combination of hyperparameters that yields the best performance on the validation set according to the specified scoring metric.
- **Final Model Training:** Once the best combination of hyperparameters is determined, it retrains the model on the entire training dataset using these hyperparameters.
- **Final Evaluation:** Optionally, it evaluates the final model on a separate holdout test set if provided, or it returns the performance metrics obtained during cross-validation.

For this study, a 3-fold cross validation and the “r^2” and “accuracy” score metrics were used for regression and classification methods respectively. The following is a brief description of the implementation of the different machine learning algorithms and the parameter grid used for the selection of the best hyperparameters.

3.2.3 Bagging

Following Breiman (1996), Bagging (Bootstrap Aggregation) consists of repeatedly taking B samples with replacement of a set of covariates X and a dependent variable Y_i, constructing trees from each sample, and subsequently combining the predictions from each tree to obtain the final prediction. Chan & Mátyás (2022) summarized the algorithm for Bagging using regression trees as:

- Draw a random sample (with replacement) of size n.
- Build a tree on each sample and obtain a prediction for a given X, $\widehat{Y_b} = \widehat{f_b}(X)$.
- Compute the average of all B predictions to get the final bagging prediction:

$$\hat{Y} = \hat{f}_{bagging}(X) = \frac{1}{B}\sum_{b=1}^{B} \widehat{f_b}(X).$$

This bagging procedure was done three times where the dependent variables for each bagging model where the years of schooling (Y_i), the teenage pregnancy variable (T_i) and the age of first menarche (Z_i) with all three of them having the same set of covariates X. For the hyperparameter tuning procedure, Probst et al. (2019) suggest that in empirical applications, 100 estimators or trees as an optimal number (B). Similarly, they suggest a value of 5 for the minimum number of observations in leaves. Moreover, Baita et al. (2023) suggest values less than 10 for the maximum depth of trees and the minimum number of observations in a sample required to split the internal node. For this, the parameter grid is defined as:

- Number of estimators: {10,30,50,100,200}
- Max depth: {1,5,10,15,20}
- Minimum sample split : {2,5,10}
- Minimum sample leaf : {1,2,4}

where "Number of estimators" is the number of trees in the bagging trees, "Max depth" is the maximum depth of the tree, Minimum sample split is the minimum number of samples required to split an internal node and Minimum sample leaf is the minimum number of samples required to be at a leaf node.

3.2.4 Random Forests

Similarly, Breiman (2001) state that the bagging method is extended by the random forests concept, which restricts the number of covariates in each bootstrapped sample. Put differently, each tree is constructed using a subset of all covariates. Generally, only $\sqrt{p}$ covariates are randomly chosen for each bootstrapped sample if there are a total of p covariates. Like Bagging, Chan & Mátyás (2022) summarized the algorithm for Random Forests using regression trees as:

- Draw a random sample (with replacement) of size n.
- Randomly select m covariates from the full set of p covariates, where $m = \sqrt{p}$.
- Build a tree using the selected m covariates and obtain a prediction for a given X, $\widehat{f_{b,m}}(x)$. This is the prediction of the bth tree based on m covariates.
- Repeat steps 1 to 3 for all B bootstrapped samples.
- Compute the average of individual tree predictions to obtain the random forest prediction:

$$\hat{f}_{rf}(x) = \frac{1}{B}\sum_{b=1}^{B}\widehat{f_{b,m}}(X).$$

For this study, the case where $m = \sqrt{p}$ was considered for both regression and classification methods. The parameter grid for Random Forest is the same as for Bagging.

3.2.5 Support Vector Regression (SVR)

Chih-Chung & Chih-Jen (2011) described the Support Vector Regression model considering a set of training points, {($\boldsymbol{x_1}$, y_1), ($\boldsymbol{x_2}$, y_2), … ,(($\boldsymbol{x_n}$, y_n))}, where $\boldsymbol{x_i} \in R^{\boldsymbol{n}}$ is a vector of features or covariates and $\boldsymbol{y_i} \in R^{\boldsymbol{1}}$ is the target output or dependent variable. For parameters $C > 0$ and $\varepsilon > 0$, the standard form of the SVR provided by (Vapnik, 1998) is:

$$
\begin{aligned}
min_{\boldsymbol{w},b,\xi,\xi^*} \quad & \frac{1}{2}\boldsymbol{w}^T\boldsymbol{w} + C\sum_{i=1}^{n}\xi_i + C\sum_{i=1}^{n}{\xi^*}_i \\
subject\ to \quad & \boldsymbol{w}^T\phi(\boldsymbol{x_i}) + b - y_i \leq \varepsilon + \xi_i \\
& y_i - \boldsymbol{w}^T\phi(\boldsymbol{x_i}) - b \leq \varepsilon + {\xi^*}_i \\
& \xi_i, {\xi^*}_i \geq 0, \qquad i = 1, 2, \dots, n.
\end{aligned}
$$

where $\boldsymbol{w}$ and b are the weights of the covariates and the intercept respectively of the hyperplane, C is the regularization constant that controls the penalty imposed on observations that lie outside the epsilon margin ε. ξ_i and ${\xi^*}_i$ are slack variables that are introduced to deal with cases which the constrains are not satisfied and $\phi(\boldsymbol{x_i})$ can be a linear or nonlinear basis function of the covariates.

The dual problem is stated as:

$$
\begin{aligned}
min_{\boldsymbol{\alpha},\boldsymbol{\alpha}^*} \quad & \frac{1}{2}(\boldsymbol{\alpha} - \boldsymbol{\alpha}^*)^T Q(\boldsymbol{\alpha} - \boldsymbol{\alpha}^*) + \varepsilon\sum_{i=1}^{n}(\alpha_i - {\alpha_i}^*) + y_i\sum_{i=1}^{n}(\alpha_i - {\alpha_i}^*) \\
subject\ to \quad & \boldsymbol{e}^T(\boldsymbol{\alpha} - \boldsymbol{\alpha}^*) = 0, \\
& 0 \leq \alpha_i, {\alpha_i}^* \leq C, i = 1, \dots, n
\end{aligned}
$$

where $\boldsymbol{e}^T = [1, \dots, 1]^T$is a vector of ones, α_i and ${\alpha_i}^*$ are non-negative multipliers and $Q_{ij} = K(\boldsymbol{x_i}, \boldsymbol{x_j}) \equiv \phi(\boldsymbol{x_i})^T\phi(\boldsymbol{x_i})$ is the kernel function. After solving the dual problem, the approximate prediction function is:

$$
\sum_{i=1}^{n}(\alpha_i - {\alpha_i}^*)\, K(\boldsymbol{x_i}, \boldsymbol{x}) + b
$$

Based on the previous framework, the first step is to choose the appropriate kernel function. Chih-Wei et al. (2003) articulate that the Radial Basis Function (RBF) kernel is a good choice of kernel where nonlinearities arise between the covariates and the dependent variable, moreover, it has fewer hyperparameters and numerical difficulties than linear kernels and performs better than linear kernels where the number of features is not very large.

The RBF kernel has the “gamma” parameter which determines how much the effect of one training example might have. This parameter can be seen as the reciprocal of the samples' radius of effect that the model used to identify support vectors and alongside with the C regularization parameter are the two main parameters for tunning. The epsilon parameter is the epsilon-tube within which no penalty associated in the training loss function with points predicted within a distance epsilon from the actual value. Chih-Wei et al. (2003) also provide a general parameter grid to test[2]:

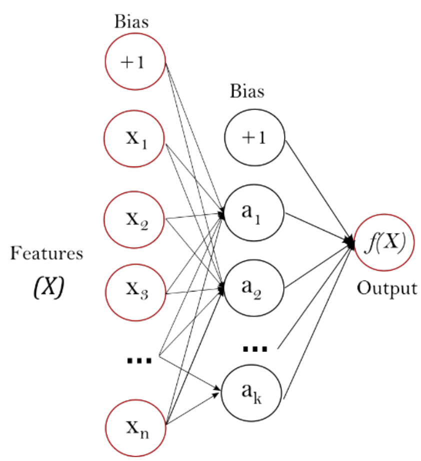
- C : {2^-5,2^-3, 2^-1,2^1, 2^3, 2^5, 2^7, 2^9}
- gamma : {2^-11, 2^-9, 2^-7,2^-5, 2^-3, 2^-1, 2^1, 2^3, 2^5}
- epsilon : {0.0001, 0.001, 0.01, 0.1}

[2] For the Support Vector Classifier, the epsilon hyperparameter is not required, so the parameter grid only considers C and gamma.

3.2.6 Multilayer Perceptron

The last ML model to consider for this methodology is the Multi-layer Perceptron (MLP). The MLP is a generalization of the "perceptron" proposed by Rosenblatt (1958) that arose from the development of the "backpropagation" algorithm popularized by Rumelhart et al. (1986). Scikit-Learn (n.d.) provide a summary of the framework for the MLP algorithm that learns from a function $f(\cdot): R^p \rightarrow R^m$ where p is the number of covariates and m is the number of dimensions for the output variable. Given a set of features $X = x_1, x_2, \ldots, x_n$ and a target y, the algorithm can learn using non-linear function approximators for classification and regression. The basic architecture of a MLP can be observed in Figure 3.

Figure 3. Basic Architecture of the MLP with one hidden layer.



The first layer, also known as the input layer, consists of a set of neurons that represent the input features. Every neuron in the hidden layer is a transformation of the of the previous values using a weighted linear summation $w_1x_1 + w_2x_2 + w_3x_3 + \cdots + w_nx_n$, followed by a non-linear activation function $f(\cdot): R \rightarrow R$. Each layer has a k number of connections and the number layers $l \geq 1$.

Heaton (2017) suggested that two or fewer layers are sufficient for simple datasets, specifically, two hidden layers can represent an arbitrary decision boundary to arbitrary accuracy with rational activation functions for classification methods and can approximate any smooth mapping to an accuracy for regression methods due to the Universal

Approximation Theory. Furthermore, he proposes a few rules of thumb for determining the number the number of acceptable neurons in each hidden layer as:

- The number of neurons should be between the input and output size layer.
- The number of neurons should be 2/3 the input size layer plus the output size layer.
- The number of neurons should be less than twice the size of the input layer.

Moreover, Jagtap & Karniadakis (2023) states thar the Rectified Linear Unit (ReLu) function is a state-of-the-art activation function that is widely used in many MLP empirical work, mainly because it is a function that ensures the Universal Approximation Theorem, it is computationally cheap, and it solves the vanishing and exploding gradient problem. This function is described as the following:

$$f(x)\colon \max(0, x).$$

Finally, Scikit-Learn (n.d.) also suggest that the “L-BFGS” optimization method converges faster with better solutions for small datasets. Based on the guidelines previously stated, the proposed neural network architecture consists of two hidden layers, where each has a ReLu activation function with the output layer having a linear activation function and the parameter grid is:

- hidden layer size : {(2,2),(3,3),(4,4),(5,5),(6,6),(7,7,)(8,8)}
- alpha: {0.00001,0.0001,0.001,0.01,0.1}

where the hidden layer size is the number of neurons in each layer and alpha is the L2 regularization term. Each pair on the parameter grid represent the number of neurons in each hidden layer. Since the MLP is a type of Neural Network, these terms are interchangeable.

Finally, although the main objective of the models is to predict continuous/categorical values for this research, each of them does it in different ways. Tree-based models are ensemble methods that base their predictions on nested decision rules for certain features or covariates. On the other hand, SVR and MLP based their predictions on model previously defined. However, SVR uses a hyperplane in a high-dimensional space to predict continuous values based on the position relative to the hyperplane while MLP uses multiple composite perceptrons or multiple layers of neurons to predict values using learned weights and activation functions.

3.2.7 Cross-Fitting

After determining the optimal hyparameters, the next and final step is to use cross-fitting for addressing the overfitting bias. For this, Chernozhukov, et al. (2018) proposed an algorithm that can be summarized as:

- Providing data as $(W_i)_{n=1}^{N}$, pick a K-fold random partition $(I_k)_{k=1}^{K}$of observations with indices $[N] = \{1, \dots, N\}$ such that the size of each fold I_k is $n = N/K$. For each $k \in [K] = \{1, \dots, k\}$, construct a ML estimator:

$$\widehat{\eta_{0,k}} = \widehat{\eta_{0,k}}((W_i)_{i \notin I_k})$$

of η_0, where $x \to \widehat{\eta_{0,k}}(x)$ depends only on the subset of data $(W_i)_{i \notin I_k}$

- Estimate the causal parameter $\widetilde{\theta_0}$ as the solution of the equation:

$$\frac{1}{N} \sum_{k=1}^{K} \sum_{i \in I_k} \psi(W_i; \widetilde{\theta_0}, \widehat{\eta_{0,k}}) = 0$$

Under regularity conditions, the estimator is $\widetilde{\theta_0}$ is approximately normal

$$\sqrt{N}\big(\widetilde{\theta_0} - \theta_0\big) \to N(0, \sigma^2),$$

with mean zero and variance given by:

$$\sigma^2 := {J_0}^{-2} E(\psi^2(W; \theta_0, \eta_0))$$

where $J_0 = E(\psi_a(W; \eta_0))$. Estimates of the variance are obtained as:

$$\hat{\sigma}^2 = \widehat{J_0}^{-2} \frac{1}{N} \sum_{k=1}^{K} \sum_{i \in I_k} \left[\psi\big(W_i; \widetilde{\theta_0}, \widehat{\eta_{0,k}}\big)\right]^2$$

$$\widehat{J_0} = \frac{1}{N} \sum_{k=1}^{K} \sum_{i \in I_k} \psi_a\big(W_i; \widehat{\eta_{0,k}}\big)$$

The cross-fitting procedure can be repeated $M > 1$ times to obtain more robust estimators to outliers. The aggregated estimators of the casual parameter are obtained using the median function as:

$$\widetilde{\theta_0} = Median\left((\widetilde{\theta_{0,m}})_{m\epsilon[M]}\right),$$

$$\hat{\sigma} = \sqrt{Median(\left(\hat{\sigma}^2{}_m + \left(\widetilde{\theta_{0,m}} - \widetilde{\theta_0}\right)^2\right)_{m\epsilon[M]})}$$

For this study, K = 3, M = 100 and the “partialling out” type score function were used following the Double ML methodology implemented in the Python package called “Double ML” from Bach et al. (2022).

4 Data

The data was retrieved from the “Encuesta Nacional de Salud y Nutricion - 2018” (ENSANUT - 2018). This is a National Survey conducted in Ecuador in 2018 whose main objective was to develop indicators on the main health and nutritional conditions and problems of the Ecuadorian population to evaluate and design public policies that properly address them. Generally, this survey was intended for women between 10 and 49 years old, kids under 5 years old, men from 12 years old and older and kids between 5 and 11 years old (Instituto Nacional de Estadisticas y Censos, 2018).

The ENSANUT 2018 is composed of 5 forms:

- Form 1: Household Members (includes anthropometry).
- Form 2: Women at Childbearing Age, Childhood Health and Breastfeeding.
- Form 3: Sexual and Reproductive Health of men from 12 years old and older.
- Form 4: Risk Factors for kids from 5 to 18 years old.
- Form 5: Child Development for children under 5 years old.

Based on the objectives established for this work, the only sections of the survey that were used were forms 1 and 2.

4.1 Stylized facts

Descriptive statistics on teenage and non-teenage pregnancy show a marked difference that could be supported by a greater knowledge of sexuality. Figure A.1[3] shows the distribution of the age interval between first sexual intercourse and first pregnancy for teenage mothers and adult mothers. It is interesting to note that 72% of teenage mothers got pregnant in the same year or the first year after the first sexual intercourse. On the other hand, adult mothers seem to have a more uniform distribution except for the case they got pregnant in the same year of their first sexual relationship.

One possible explanation for this difference may be related the usage and type of contraceptive. Table 1 shows that for teenage mothers, 31% assured to have use a

[3] All figures that start with “A.” are in the Figures subsection of the Annexes. Likewise, all tables that start with “A.” are in the Tables subsection of the Annexes.

contraceptive method in contrast to the 42% for adult mothers. Although there is a significant difference between the two groups, it is interesting to note that even for adult mothers, most women did not use any type of contraceptive method.

Table 1. Contraceptive methods in first sexual relationship

Did you use a contraceptive method in your first sexual relationship?	Teen Pregnancy		
	No	Yes	Total
Yes	41.67	31.23	33.14
No	58.33	68.77	66.86
Total	100.00	100.00	100.00

Figure A.2 shows the different reasons why mothers did not decide to use any contraceptive method. Both group of mothers have a similar behavior towards the reasons for not using contraceptive methods in their first sexual relationships. Although most mothers had information about the different contraceptive methods, the main reason for not using them was the fact that they did not expect to have sexual intercourse at the time. This suggests a sexual behavior more related to recreational than reproductive purposes.

On the other hand, Figure A.3 shows the different contraceptive methods used by women. Despite the availability of several options in the Ecuadorian market, 76% of adolescents and non-adolescents decided to use condoms, which could mean an aversion to the risk of contracting sexually transmitted diseases, higher costs of complementary methods, or low availability of these methods. The low intake of contraceptive pills (8% and 7% respectively), as well as low participation of other methods such as contraceptive vaccines could be associated with the lack of sexual transmitted disease protection.

A topic that causes national and international debate is the partner's age with whom a woman had her first sexual relationship. Figure A.4 shows that most women had sexual partners with ages between 20 and 40 years old. This could be due to rational choice for young women facing poverty and lack of opportunities, as they have the chance to escape from domestic violence, have control over their lives or even access social mobility (World Bank, 2015).

Unfortunately, a poignant fact from Figure A.4 shows that all minor girls have suffered sexual abuse from men between 18 to 65 years old, with the distressing fact that girls between 10 to 15 years old suffered sexual abuse from people with 30 years old or above. A possible explanation for this can be found in a study developed by Casas et al. (2016). They analyzed the clinical records from pregnant girls between 10 to 14 years old and found out that in Ecuador, in 51% of the cases, the newborn's fathers are of legal age. Moreover, 12% of teenagers reported pregnancies resulting from sexual abuse, 44% of which are sexual abuse or rape committed by family members. Finally, 82% are unwanted/unplanned pregnancies.

Although girls are vulnerable to sexual abuse and unwanted pregnancies as kids and teenagers, this trend seems to be changing over time. Figure A.5 shows that for adult mothers, there is a reduction in the variability and amount of father's age data when controlling for each year. This reduction may be related to the fact that more education and life experience makes them feel less threatened by lack of opportunities and less likely to be victims of sexual abuse or non-consensual sex.

To have a first look at the educational level of Ecuadorian mothers, it is necessary to know about the situation and the decisions they made when facing pregnancy. First, Table 2 shows that 50% of women that got pregnant where studying, 49% were not studying and less than 1% have never studied. The fact that only 50% of pregnant women were attending school at the time they became pregnant suggests a low schooling rate among the women surveyed.

Table 2. Distribution of women that got pregnant while studying

Were you a student when you knew you were pregnant?	Percent
Yes	50.65
No	49.20
I have never studied	0.15
Total	100.00

Once women are pregnant, they must take the difficult decision between interrupting their studies or continuing them and then, once the pregnancy is over, considering whether it is suitable for them to resume their studies. For Ecuadorian women, Table 3 shows that 52% of woman had to interrupt their studies while being pregnant. Of these women, 35% did not resume their pregnancies. In contrast, for the 47% of woman that did not interrupt their studies, 83% of them keep studying. The 17% left of woman who did not interrupt their studies and yet did not resume them may be related to cases although they might have finished the schooling year and chose not to study afterwards for economic, parental, social reasons, etc.

Table 3. Distribution of women that interrupted and resumed their studies after pregnancy

Did you interrupt your studies because of your pregnancy?	Did you resume your studies after the pregnancy ended?		
	Yes	No	Total
Yes	17.99	34.41	52.41
No	39.41	8.18	47.59
Total	57.41	42.59	100.00

The number of women who do not return to school after pregnancy reaches an alarming 42%. Although there are many reasons for women not to resume their studies after pregnancy, the main ones for Ecuadorian woman are because no one else could take care of the baby and she had no time left as showed in Table 4. These reasons mainly show that although most women want to resume their studies, the responsibility of having a child is too great to continue studying and pursuing their academic and professional dreams. This situation could be lessened with government investment to facilitate women's access to education, such as free day care centers, more institutions that provide evening classes, among others.

Table 4. Reasons for not resuming your studies

Why didn't you continue studying after pregnancy?	Percent	Cum.
No one could take care of my baby	54.14	54.14
I had no time left	22.97	77.10
Other	12.33	89.44
Changed my address	2.22	91.65
My partner did not want me to study	2.14	93.80
Health problems	2.07	95.86
Had to work	2.07	97.93
The baby was sick	1.62	99.56
Was not admitted/ was expelled	0.37	99.93
Sexual Harassment	0.07	100.00
Total	100.00	

4.2 Variables

Although there is abundant economic literature relating teenage pregnancy and women's years of schooling, the selected covariates and instrument were chosen based on their availability in the ENSANUT - 2018. Table 5 shows a list of all the variables selected for the different models with their respective definitions.

Table 5. List of models' variables with their definitions

Variable's names	Definition
years_educ	Schooling years (Y_i)
teen_pregnancy	Dummy for teen pregnancy (T_i)
menarche	Age of menarche (Z_i)
p_age_bfsr	Partner's age before first sexual relationship
contraceptive	Dummy for using a contraceptive method at first sexual relationship
highlands_region	Dummy if the woman lives in the highland's region
coastal_region	Dummy if the woman lives in the coastal region
amazon_region	Dummy if the woman lives in the amazon region
afro_equadorian	Dummy if the woman identify herself as afro-ecuadorian
mixed	Dummy if the woman identify herself as mixed
white	Dummy if the woman identify herself as white
urban	Dummy if the woman lives in an urban area
fathers_educ	Father's schooling years
mothers_educ	Mother's schooling years
siblings	Number of Siblings

Although the number of father's and mother's schooling years are ubiquitous variables in educational studies, and to a lesser extent, the number of siblings, they are not considered for the main model but rather for a secondary model, the results of which are presented in the annexes. The main reason for this is because these variables are only obtainable by reducing the number of observations because the primary observational unit were households and their members. Since most of the households are constituted by a father, mother, and their children, most of the individuals from whom information on the years of schooling of their parents can be obtained are the children of the household, eliminating potential observations such as mothers, aunts, grandmothers, or other female members of the household. This issue reduced the sample size from 6025 to 1172 individuals. Because of the considerable reduction in the sample size, the main model will consider previous variables without the father's and mother's schooling years and the number of siblings, and the secondary model will consider them all.

4.3 Descriptive Statistics

The univariate descriptive statistic for the variables is presented in Table 6 which suggests interesting results. First, 81% of mothers had a teenage pregnancy. Second, the average years of schooling for mother is 11.29 years which is related for students that did not finish their secondary education.

Moreover, the average girl had her first menarche at the age of 12 and only 33% of them used contraceptive methods in their first sexual encounter. The average age of their partner's first sexual relationship is 20 years although there are cases of girls having this experience with adults up until the age of 60 years old. Finally, the mothers are similarly distributed among the different regions, 75% of them perceived themselves as mixed and 57% live in urban areas.

Table 6. Univariate descriptive statistics for main model

	Mean	Med.	S.D.	Min	Max	p25	p75	Skew	Kurt
years_educ	11.29	13	3.010	0	20	10	13	-.9	3.94
teen_pregnancy	.82	1	0.390	0	1	1	1	-1.64	3.69
menarche	12.64	12	1.430	8	19	12	14	.32	3.41
p_age_bfsr	20.17	19	4.270	12	60	18	22	2.24	12.5
contraceptive	.34	0	0.470	0	1	0	1	.69	1.48
highlands_region	.37	0	0.480	0	1	0	1	.56	1.31
coastal_region	.39	0	0.490	0	1	0	1	.44	1.19
amazon_region	.23	0	0.420	0	1	0	0	1.31	2.71
afro_ecuadorian	.05	0	0.210	0	1	0	0	4.27	19.26
mixed	.75	1	0.430	0	1	1	1	-1.16	2.35
white	.01	0	0.110	0	1	0	0	9.18	85.33
urban	.58	1	0.490	0	1	0	1	-.32	1.1

Although the previous descriptive statistics provide a general perspective of the sample, it is also interesting to investigate on the general picture of both teenage and non-teenage mothers. Tables 7 and 8 present the descriptive statistics of the main variables, excluding demographic variable for teenage and non-teenage mothers respectively. These tables show that there are important differences between the number of average schooling years, the partner's age before the first sexual relationship and the percentage of woman that used contraceptive methods in their first sexual encounter but there also might be an insignificant difference between the age of menarche.

Table 7. Univariate descriptive statistics for teenage mothers

	Mean	Med.	S.D.	Min	Max	p25	p75	Skew	Kurt
years_educ	11.01	12	2.960	0	20	9	13	-.93	3.86
menarche	12.59	12	1.410	8	19	12	13	.34	3.47
p_age_bfsr	19.96	19	4.210	12	60	17	22	2.18	11.9
contraceptive	.32	0	0.470	0	1	0	1	.79	1.62

Note: *years_educ* is the schooling years, *menarche* is the age the first menarche, *p_age_bfsr* is the partner's age before the first sexual relationship and *contraceptive* is an indicator variable for the case where the individual used a contraceptive method in the first sexual relationship.

Table 8. Univariate descriptive statistics for non-teenage mothers

	Mean	Med.	S.D.	Min	Max	p25	p75	Skew	Kurt
years_educ	12.58	13	2.860	1	20	13	13	-1.09	5.04
menarche	12.86	13	1.500	9	18	12	14	.18	3.23
p_age_bfsr	21.09	20	4.430	12	58	18	23	2.56	14.9
contraceptive	.42	0	0.490	0	1	0	1	.3	1.09

Note: *years_educ* is the schooling years, *menarche* is the age the first menarche, *p_age_bfsr* is the partner's age before the first sexual relationship and *contraceptive* is an indicator variable for the case where the individual used a contraceptive method in the first sexual relationship.

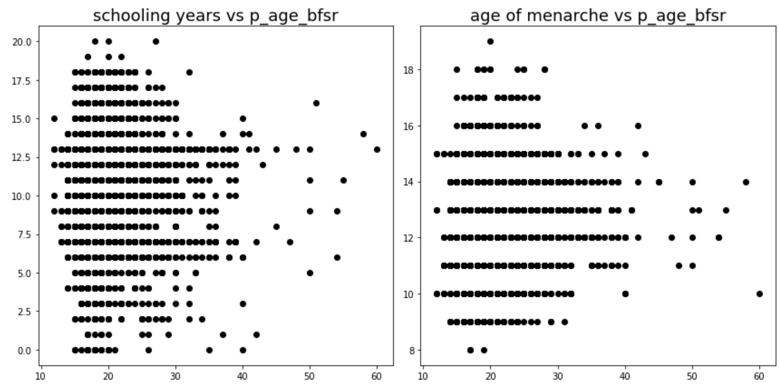
Table 9 show the correlations of the three previous variables with the only continuous covariate[4] where *years_educ* is the schooling years, *teen_pregnancy* is the indicator variable for teenage pregnancy, *menarche* is the age the first menarche, *p_age_bfsr* is the partner's age before the first sexual relationship. The fact that there are small values for the correlations might an indicative of non-linear relationships or associations among the data. Furthermore, possible non-linear relationships can be observed in Figure 4.

Table 9. Correlations between dependent, endogenous and instrument variables with continuous covariates

Variables	(1)	(2)	(3)
(1) years_educ	1.000		
(2) teen_pregnancy	-0.203	1.000	
(3) menarche	0.029	-0.073	1.000
(4) p_age_bfsr	-0.044	-0.102	0.016

[4] Since the rest of the covariates are dummy variables, non-linear relationships are assumed.

Figure 4. Scatterplots of the dependent and IV with partner's age before first sexual relationship



A similar analysis can be done with the inclusion of variables like the father's and mother's schooling years and the number of siblings at the expense of the reduction in the number of observations in the sample. Table A.1 shows the main the descriptive statistics of the new sample, having similar descriptive statistics compared to the ones with the main sample with the addition that the average number of schooling years of fathers and mother is less than the average number of schooling years of the individuals and the individuals have on average three siblings.

Similarly, Tables A.2 and A.3 also present a similar behavior like the main sample, specifically, difference among the number of schooling years, partner's age before the first sexual relationship and the percentage of people that used a contraceptive method in their first sexual relationship. Moreover, there seems to be insignificant differences among the rest of the variables between teenagers and non-teenager mothers. Finally, Table A.4 also show small values for correlations between the dependent, endogenous and instrument variables with the covariates except for the case of the correlations of the father's and mother's education and the number of siblings. Although these values are bigger that the rest of the correlations, they still imply a no correlation or very weak correlation. Moreover, Figure A.6 shows scatterplots of the dependent and IV with every continuous covariate where linear relationships appear not to be present.

5 Results

The first step to properly integrate the IV with the Double/Debiased Machine framework is to determine if the selected instrument is a valid instrument and satisfies the relevance and exogeneity assumptions and if the teen pregnancy variable is endogenous. For the Relevance condition, the Montiel-Pflueger robust weak instrument test was performed yielding the following results:

Table 10. Robust Weak IV test

Montiel-Plueger robust weak instrument test		
Effective F-statistic:	32.483	
Confidence level alpha:	5%	
Critical Values	TSLS	LIML
% of Worst Case Bias		
τ = 5%	37.418	37.418
τ = 10%	23.109	23.109
τ = 20%	15.062	15.062
τ = 30%	12.039	12.039

Table 10 reports that the F-statistic for the first stage of the TSLS estimator. The null hypothesis for each level of τ is that the estimator's approximate asymptotic bias exceeds the fraction tau of a "worst-case" benchmark. The test rejects the null hypothesis when the Effective F-statistic exceed the critical value. For our scenario, since the Effective F-statistic is greater than critical value for a 10% worst case bias with a 5% significance level, the instrument is not weak. The 10% worst case bias is the benchmark used by Montiel & Pflueger (2013).

Results for the endogeneity tests are presented in Table 11. Both tests reject the null hypothesis that the variable related to teen pregnancies are exogenous suggesting it is an endogenous variable.

Table 11. Endogeneity tests

Tests for Endogeneity		
H0: Variables are exogenous	Statistic	P-value
Robust score chi2(1)	7.18339	0.0074
Robust regression F(1,6013)	7.17756	0.0074

After determining that the age of menarche in girls is a good instrument, the next step is to determine the best hyperparameter for every ML model for the schooling years, teen pregnancy and age of menarche and covariates. For this, a hyperparameter tunning procedure was used yielding the following results in Table 12, 13 and 14. Results of the hyperparameter tunning procedure for the secondary model can be found in Table A.5, A.6 and A.7.

Table 12. Hyperparameter results for menarche and covariates.

Hyperparameter Tunning Results							
Random Forest		**Bagging**		**SV Regression**		**Neural Networks**	
n_estimators	50	n_estimators	50	C	0.5	hl_size	(7,7)
max_depth	5	max_depth	1	gamma	0.125	alpha	10
min_s_split	1	min_s_split	1	epsilon	0.1		
min_s_leaf	1	min_s_leaf	1				

Table 13. Hyperparameter results for teen pregnancy and covariates.

Hyperparameter Tunning Results							
Random Forest		**Bagging**		**SV Classifier**		**Neural Networks**	
n_estimators	10	n_estimators	10	C	0.0312	hl_size	(6,6)
max_depth	1	max_depth	1	gamma	2^(-11)	alpha	10
min_s_split	1	min_s_split	2				
min_s_leaf	1	min_s_leaf	1				

Table 14. Hyperparameter results for years of education and covariates.

Hyperparameter Tunning Results							
Random Forest		**Bagging**		**SV Regression**		**Neural Networks**	
n_estimators	200	n_estimators	100	C	128	hl_size	(5,5)
max_depth	1	max_depth	1	gamma	2^(-9)	alpha	10
min_s_split	1	min_s_split	1	epsilon	0.1		
min_s_leaf	1	min_s_leaf	1				

Finally, after tunning the hyperparameters, the Double/Debiased ML methodology was implemented yielding the results in Table 15[5] where RF, BAG, SVR and NN stand for Random Forest, Bagging, Support Vector Regression and Neural Network regression

[5] The Double ML package in python does not allow for the integration of HAC Robust Standard Errors because the methodology assumes homoscedastic errors. Because of this, the Hall & Pagan heteroscedasticity test for IV was performed showing the presence of heteroscedasticity. But, after estimating the causal parameter using 2SLS with and without Heteroscedasticity Robust Standard Errors, the difference between them is negligible. This may indicate that Heteroscedasticity may not influence the statistical significance for all models, or at least, for the model that uses OLS.

respectively as the ML models. For the OLS model, which is the same as the TSLS estimator and is the benchmark for this study, suggest that teenage mothers on average have 4.70 less schooling years than non-teenage mothers. Similarly, the Tree-based methods such as RF and Bagging provide similar results like 4.34 and 4.28 less schooling years respectively with a little difference with respect to the benchmark method. Similarly, the MLP model suggest that teenage mothers on average have 4.35 less schooling years than non-teenage mothers.

These results are consistent with the negative effect of teenage pregnancy on women's years of schooling, but they differ in the magnitude. A possible explanation for the substantial discrepancy from prior findings may be attributed to the inherent differences in quality of life as well as the different mindsets and tools people from different countries possess to face this problem. Likewise, another plausible reason for these differences might be the different methodologies and the way in which the years of schooling are defined.

Finally, the SVR method provides an estimate completely unrelated to the others while being statistically not significant. Chih-Wei et al. (2003) argue about the importance of feature scaling in the correct performance of the algorithm. Moreover, Xulei et al. (2007) state that outliers also affect the performance of the model since outliers can become the support vectors while traning the model, leading to decision boundaries that severely deviate from the optimal hyperplane. The data is not scaled to keep interpretability for the DML resutls and Figure 4 shows the possible prescence of outliers for the partner's age before the first sexual relationship. These circumstances might be the reason behind the different resutls obtained for SVR models, unfortunately, it is not possible to know for certain.

Table 15. Double ML Results

	OLS	**RF**	**BAG**	**SVR**	**MLP**
coef	-4.702	-4.349	-4.289	-0.478	-4.356
std. error	1.355	1.492	1.540	0.834	1.512
T	-3.469	-2.915	-2.785	-0.573	-2.882
P>|t|	0.001	0.004	0.005	0.567	0.004
2.5 %	-7.358	-7.273	-7.307	-2.113	-7.319
97.5 %	-2.045	-1.424	-1.271	1.157	-1.393

Note: OLS Standard-Errors (Non-robust to heteroscedasticity)

Similarly, results for the secondary model can be found in Table A.8. The inclusion of the variables related to the father's and mother's education and the number of siblings decreased the estimates the results. However, for this case, the results follow a different pattern. First, the results obtained are lower in magnitude than those compared with OLS. Second, results for all ML methods are not statistically significant at 5% except for the NN model.

Unfortunately, as in the general model, the underlying reasons for the inconsistent estimations for the Double ML with Support Vector Machines might be related to data scaling and outliers. For this study, results suggest the use of Tree-based methods and Multi-Layer Perceptron are good methods to account for the non-linearities in data and providing less biased estimators for the LATE. Additionally, it seems that using SVR with DML does not produce consistent results.

6 Conclusions

Teenage pregnancy is a multidimensional problem that affects many women around the world and has serious consequences in different aspects of their lives, with Latin America being one of the most affected regions. Ecuador is one of the countries that suffers the most from this problem, having one of the highest teenage pregnancy rates in the region. Although there are studies on its consequences, most of them are incipient and focus mainly on correlational studies. Therefore, the objective of the present study focuses on determining the causal effect of pregnancy on the years of schooling of adolescent girls.

By combining the Double/Debiased Machine Learning methodology alongside the IV framework, this study provides evidence of the negative effect of teenage childbearing in women in Ecuador. Results suggests that on average, teenage mothers have four less schooling years than non-teenage mothers[6]. These results are aligned with the negative results provided by previous literature but the significant greater difference with previous results may be related to the inherent differences in living conditions between countries and the ways in which they deal with the problem, or the different methodologies used for each study.

The main contributions of this study are the empirical applications of the IV framework to study the effects of teenage pregnancy in the schooling years of woman in a country like Ecuador and the inclusion of DML with IV in this context. The results also provide empirical evidence of the causal effects in teenage pregnancy in the education of women in Ecuador. To the best of my knowledge, this is one of the few studies that addresses this problem with scientific rigor in Ecuador and the only one that uses a wide range of modern econometric methods, such as DML, that allow to flexibly allow to capture complex (non-linear) relationship among the data while at the same providing causal estimates.

The results hereby presented can be of great help mainly for government authorities and public policy makers in Ecuador. They can help them to understand the

[6] If father's education, mother's education, and the number of siblings is considered, the effect is decreased to six years less.

magnitude of this problem to reduce the negative economic and social implications of low levels of schooling. The possible economic implications may be a lower labor income, higher unemployment rates, lower labor productivity, greater economic dependence on social programs, and the social implications can be greater social inequality, greater exposure to environments with higher levels of insecurity and crime, less access to health and social welfare, among others in women.

Moreover, these results can motivate politicians and public policy makers to propose solutions such as more flexible educational programs like night classes, online courses, or diverse school schedules, provide affordable or subsidized support services like day care centers and transportation assistance or financial assistance like scholarships or grants among others. These suggestions can help to tackle the consequences of early childbearing in education, a problem that affects many adolescent girls, a problem with consequences in the short, medium, and long term which eventually inhibits them from accessing a higher level of schooling and therefore, a better quality of life, robbing them of the hope of a better future.

Despite these significant findings, several limitations must be considered. First, although the data sample was taken from a nationwide survey, the results of this study cannot be generalized for all of Ecuador because many women preferred not to answer the age of their first menarche, which significantly reduced the size of the sample. Similarly, the survey did not include ubiquitous variables found in economic literature like the father's and mother's education and in less extent, the number of siblings and they had to be calculated using the information from the survey, however, this also significantly reduced the sample size.

Future research should focus on the inclusion of more variables that can be used as instruments for teenage pregnancy for better causal estimates. Moreover, the inclusion of other tree-based methods like XG-Boost or other ML methods like K-nearest neighbors, Naïve Bayes or Deep Neural Networks could also be considered.

In conclusion, the present study seeks to demonstrate that teenage mothers have, on average, four years less of schooling that non-teenage mothers, to provide empirical evidence of this problem and to put it in the public arena so that, at least, a debate can be generated and the authorities become aware of this difficult situation and effective measures can be taken that are sustainable over time.

7 Bibliography

Ashcraft, A., & Lang, K. (2006). The consequences of teenage childbearing. *National Bureau of Economic Research, 12485*. https://doi.org/10.3386/w12485

Ashcraft, A., Fernandez-Val, I., & Lang, K. (2013). The consequences of teenage childbearing: Consistent estimates when abortion makes miscarriage nonrandom. *The Economic Journal, 123*(571), 875–905. https://doi.org/10.1111/ecoj.12005

Bach, P., Chernozhukov, V., Kurz, M., & Spindler, M. (2022). DoubleML - An object-oriented implementation of double machine learning in python. *Journal of Machine Learning Research, 23*(53), 1−6. https://doi.org/10.48550/arXiv.2104.03220

Baita, A., Prasetyo, I., & Cahyono, N. (2023). Hyperparameter tuning on random forests for diagnose COVID-19. *Jurnal Informatika dan Komputer, 6*(2), 138-143. https://doi.org/10.33387/jiko.v6i2.6389

Becker, G. (1981). *A treatise on the family.* Cambridge, MA: Harvard University Press. https://doi.org/10.4159/9780674020665

Breiman, L. (1996). Bagging predictors. *Machine Learning, 24*, 123-140. https://doi.org/10.1007/BF00058655

Breiman, L. (2001). Random forests. *Machine Learning, 45*, 5-32. https://doi.org/10.1023/A:1010933404324

Card, J., & Wise, L. (1978). Teenage mothers and teenage fathers: The impact of early childbearing on the parents' personal and professional lives. *Family Planning Perspectives, 10*(4), 199-205. https://doi.org/10.2307/2134267

Casas, X., Cabrera, O., Reingold, R., & Grossman, D. (2016). *Vidas Robadas - Un estudio multipais sobre los efectos en la salud de las maternidades forzadas en niñas de 9–14 años.* Planned Parenthood Action Fund. https://www.plannedparenthoodaction.org/uploads/filer_public/cc/16/cc169bc8-1e15-49b1-a20f-62e0dd4d6c35/ppfa-stolen-lives-spanish.pdf

Chan, F., & Mátyás, L. (2022). *Econometrics with Machine Learning.* Springer. https://doi.org/10.1007/978-3-031-15149-1

Chernozhukov, V. et al. (2018). Double/Debiased machine learning for treatment and vausal parameters. *The Econometrics Journal, 21*(1), 1-71. https://doi.org/10.48550/arXiv.1608.00060

Chevalier, A., & Viitanen, T. (2002). The causality between female labour force participation and the availability of childcare. *Applied Econometrics Letters, 9*(14), 915-918. https://doi.org/10.1080/13504850210138469

Chih-Chung, C., & Chih-Jen, L. (2011). LIBSVM: A library for support vector machines. *ACM Transactions on Intelligent Systems and Technology, 2*(3), 1-27. https://doi.org/10.1145/1961189.1961199

Chih-Wei, H., Chih-Chung, C., & Chih-Jen, L. (2003). *A Practical Guide to Support Vector Classification.* Department of Computer Science and Information Engineering, University of National Taiwan. https://www.csie.ntu.edu.tw/~cjlin/papers/guide/guide.pdf

Diaz, C., & Fiel, J. (2016). The effect(s) of teen pregnancy: Reconciling theory. *Demography, 53*(1), 85-116. https://doi.org/10.1007/s13524-015-0446-6

Fletcher, J., & Wolfe, B. (2009). Education and labor market consequences of teenage childbearing. *Journal of Human Resources, 44*(3), 303-325. https://doi.org/10.3386/w13847

Geronimus, A., & Korenman, S. (1992). The socioeconomic consequences of teen childbearing reconsidered. *The Quarterly Journal of Economics, 107*(4), 1187-1214. https://doi.org/10.2307/2118385

Grogger, J., & Bronars, S. (1993). The socioeconomic consequences of teenage childbearing: Findings from a Natural Experiment. *Family Planning Perspectives, 25*(4), 156-161+174. https://doi.org/10.2307/2135923

Hayes, C. (1987). Adolescent pregnancy and childbearing: An emerging research. In C. Hayes, *Risking the future: Adolescent sexuality, pregnancy, and childbearing* (pp. 1-6). Washington, D.C.: National Academic Press. https://doi.org/10.1002/9780470756607.ch20

Heaton, J. (2017). *The Number of Hidden Layers*. Heaton Research: https://www.heatonresearch.com/2017/06/01/hidden-layers.html

Hofferth, S. (1987). Social and economic consequences of teenage childbearing. In C. Hayes, *Risking the Future: Adolescent Sexuality, Pregnancy, and Childbearing.* Washington, D.C.: National Academy Press. https://doi.org/10.1002/9780470756607.ch20

Hoffman, S., Foster, M., & Furstenberg Jr., F. (1993). Reevaluating the costs of teenage Childbearing. *Demography, 30*(1), 1-13. https://doi.org/10.2307/2061859

Holmund, H. (2005). Estimating long-term consequences of teenage childbearing: An examination of the siblings approach. *The Journal of Human Resources, 40*(3), 716-743. https://doi.org/10.3368/jhr.XL.3.716

Imbens, G., & Wooldridge, J. (2009). Recent developments in the econometrics of program evaluation. *Journal of Economic Literature, 47*(1), 5-86. https://doi.org/10.1257/jel.47.1.5

Instituto Nacional de Estadisticas y Censos. (2018). *Encuesta Nacional de Salud y Nutricion.* Quito: Ecuador en Cifras. https://www.ecuadorencifras.gob.ec/salud-salud-reproductiva-y-nutricion/

Jagtap, A., & Karniadakis, G. (2023). How important are activation functions in regression and classification? A survey, performance comparison, and future directions. *Journal of Machine Learning for Modeling and Computing, 4*(1), 21-75. https://doi.org/10.1615/JMachLearnModelComput.2023047367

Keplinger, D., Lundberg, S., & Plotnick, R. (1999). How does adolescent fertility affect the human capital and wages of young women? *Journal of Human Resources, 34*(3), 421-448. https://doi.org/10.2307/146375

Lee, D. (2010). The early socioeconomic effects of teenage childbearing: A propensity score matching approach. *Demographic Research, 23*(25), 697-736. https://doi.org/10.4054/DemRes.2010.23.25

Levine, D., & Painter, G. (2003). The schooling costs of teenage out-of-wedlock childbearing: Analysis with a within-school propensity-score-matching estimator. *The Review of Economics and Statistics, 85*(4), 884-900. http://www.jstor.org/stable/3211812

Manlove, J. (1998). The influence of high school dropout and school disengagement on the risk of school-age pregnancy. *Journal of Research of Adolescence, 8*(2), 187-220. https://doi.org/10.1207/S15327795JRA0802_2

Moler-Zapata, S., Grieve, R., & O'Neill, S. (2023). How does a local instrumental variable method perform across settings with instruments of different strenghts? A simulation study and a evaluation of emergency surgery. *Health Economics, 32*(9). https://doi.org/10.1002/hec.4719

Montiel, J., & Pflueger, C. (2013). A robust test for weak Instruments. *Journal of Business &*

Economics Statistics, 31(3), 358-369. https://doi.org/10.1080/00401706.2013.806694
Moore, K., & Waite, L. (1977). Early childbearing and educational attainment. *Family Planning Perspectives, 9*(5), 220-225. https://doi.org/10.2307/2134432
Mott, F., & Marsiglio, W. (1985). Early childbearing and completion of high school. *Family Planning Perspectives, 17*(5), 234-237. https://doi.org/10.2307/2135098
Nagar, A. (1959). The bias and moment matrix of the general k-class estimators of the parameters in simultaneous equations. *Econometrica, 27*(4), 575-595. https://doi.org/10.2307/1909352
Newey, W., & West, K. (1987). A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica, 55*(3), 703-708. https://doi.org/10.2307/1913610
Olsen, R., & Farkas, G. (1989). Endogenous covariates in duration models and the effect of adolescent childbirth on schooling. *The Journal of Human Resources, 24*(1), 39-53. https://doi.org/10.2307/145932
Pedregosa, F. et al. (2011). Scikit-learn: Machine learning in python. *Journal of Machine Learning Research, 12*(85), 2825-2830. https://doi.org/10.48550/arXiv.1201.0490
Presser, H. (1978). Age at menarche, socio-sexual behavior, and fertility. *Biodemography and Social Biology, 25*(2), 94-101. https://doi.org/10.1080/19485565.1978.9988327
Probst, P., Wright , M., & Boulesteix, A. (2019). Hyperparameters and tuning strategies for random forest. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 9*(3). https://doi.org/10.1002/widm.1301
Ribar, D. (1994). Teenage fertility and high school completion. *Review of Economics and Statistics, 76*(3), 413-424. https://doi.org/10.2307/2109967
Ribar, D. (1999). The socioeconomic consequences of young women's childbearing: Reconciling disparate evidence. *Journal of Population Economics, 12*(4), 547-565. https://www.jstor.org/stable/20007649
Rindfuss, R., Bumpass, L., & St. John, C. (1980). Education and fertility: Implications for the roles women occupy. *American Sociological Review, 45*(3), 431-447. https://doi.org/10.2307/2095176
Rosenblatt, F. (1958). The perceptron: A probabilistic model for information storage and organization in the brain. *Psychological Review, 65*(6), 386-408. https://doi.org/10.1037/h0042519
Rumelhart, D., Hinton, G., & Williams, R. (1986). Learning representations by back-propagating errors. *Nature*(323), 533-536. https://doi.org/10.1038/323533a0
Sanders, S., & Zhang, Y. (2007). *Teenage Childbearing and Maternal Schooling Outcomes: Evidence from Matching.* College Park: Unpublised Manuscript, Department of Economics, University of Maryland.
Scikit-Learn. (n.d.). *1.17. Neural network models (supervised)*. Scikit-Learn.org: https://scikit-learn.org/stable/modules/neural_networks_supervised.html#mlp-tips
Staiger, D., & Stock, J. (1997). Instrumental variables regression with weak instruments. *Econometrica, 65*(3), 557-586. https://doi.org/10.2307/2171753
Stock, J., & Watson, M. (2018). *Introduction to Econometrics.* Pearson.
Stock, J., & Yogo, M. (2005). Testing for weak instruments in linear IV regression. In J. Stock, & D. Andrews, *Identification and Inference for Economic Models: Essays in Honor of Thomas Rothenberg* (pp. 80–108). New York: Cambdrige University Press. https://doi.org/10.1017/CBO9780511614491

UNFPA (2020). *Socioeconomic consequences of adolescent pregnancy in six Latin American countries. Implementation of the MILENA methodology in Argentina, Colombia, Ecuador, Guatemala, Mexico and Paraguay.* United Nations Population Fund - Latin America and the Caribbean Regional Office. https://lac.unfpa.org/sites/default/files/pub-pdf/eng_-_informe_consecuencias_socioeconomicas_del_embarazo_en_la_adolescencia_en_seis_paises_de_america_latina_y_el_caribe._vf.pdf

Upchurch, D., & McCarthy, J. (1990). The timing of a first birth and high school completion. *American Sociological Review, 55*(2), 224-234. https://doi.org/10.2307/2095628

Vapnik, V. (1998). *Statistical Learning Theory.* Wiley.

Waite, L., & Moore, K. (1978). The impact of an early first birth on young women's educational attainment. *Social Forces, 56*(3), 845-865. https://doi.org/10.2307/2577222

Wooldridge, J. (2002). *Introductory Econometrics: A Modern Approach.* Michigan: South-Western Cengage Learning, 2013.

World Bank (2015). *Embarazo adolescente, un escape ante la falta de oportunidades*. https://www.bancomundial.org/es/news/feature/2015/09/01/teenage-pregnancy-latin-america

World Bank (2024). *World Development Indicators*. https://data.worldbank.org/indicator/SP.ADO.TFRT

World Health Organization (n.d.). *Adolescent health*. https://www.who.int/health-topics/adolescent-health#tab=tab_1

Xulei, Y., Qing, S., & Yue, W. (2007). A weighted support vector machine for data classification. *International Journal of Pattern Recognition and Artificial Intelligence, 21*(5), 961-976. http://dx.doi.org/10.1109/IJCNN.2005.1555965

8 Annexes

8.1 Graphs

Figure A.1 Distribution of the age interval between first sexual intercourse and first pregnancy

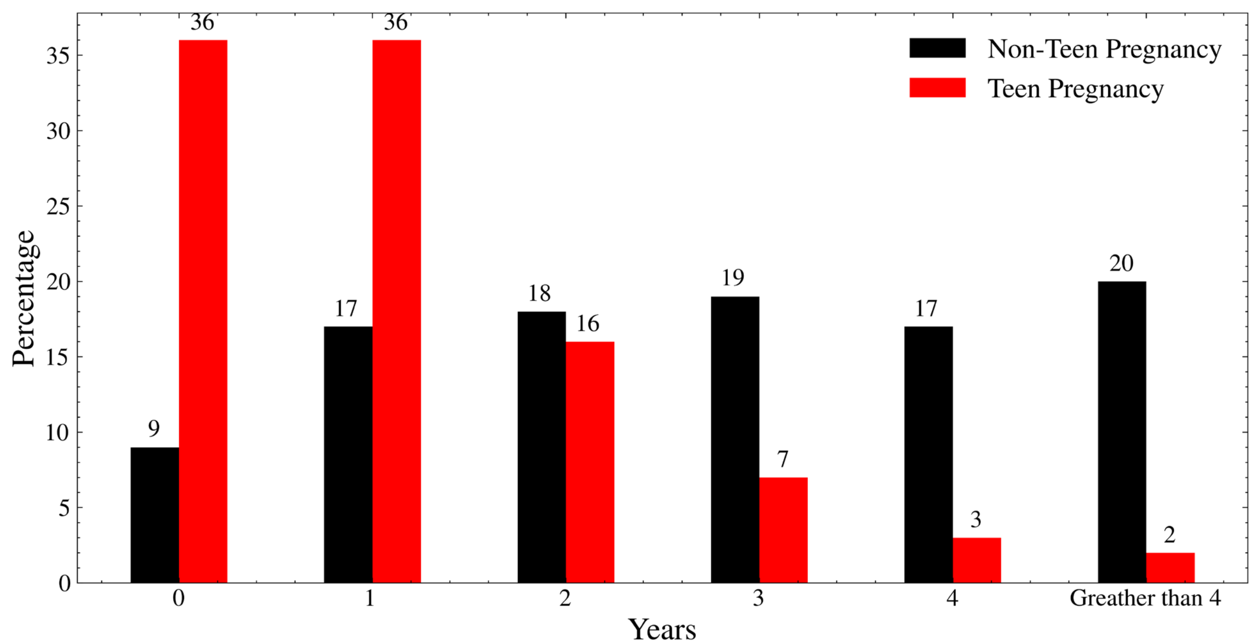


Figure A.2 Reasons why you did not use any conceptive method in your first relationship

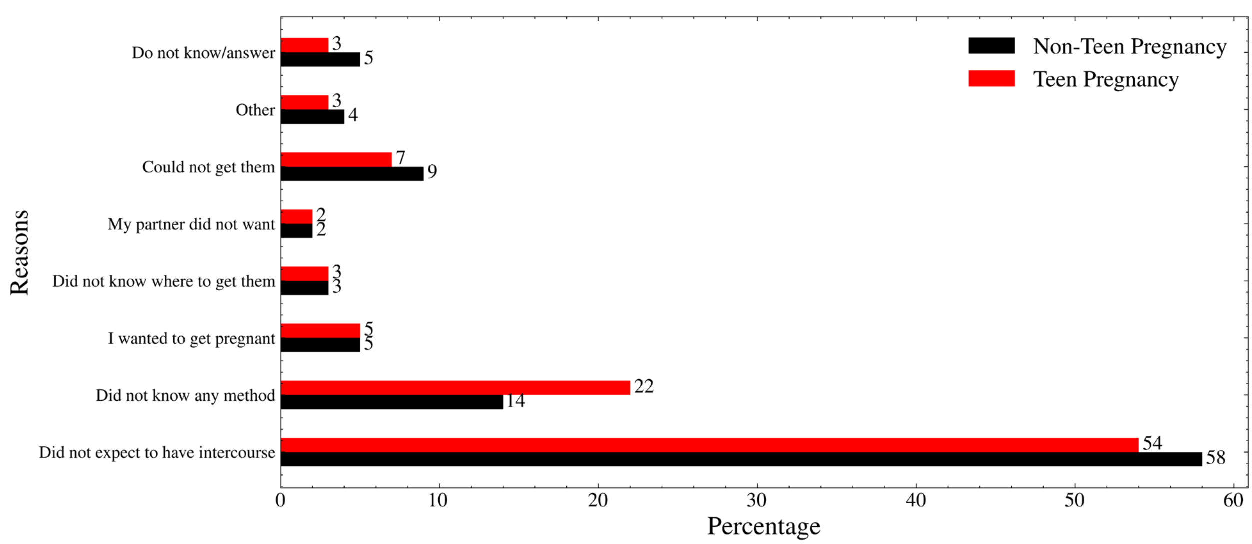


Figure A.3 Contraceptive method used in first sexual intercourse

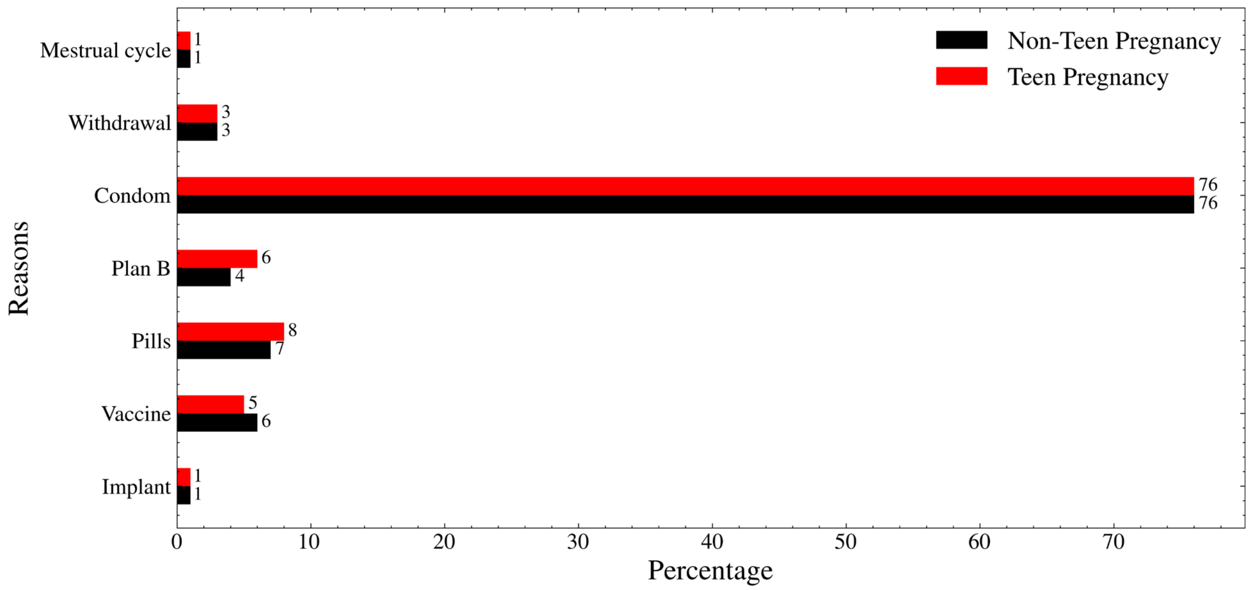


Figure A.4 Relationship between father's and mother's age of first pregnancy

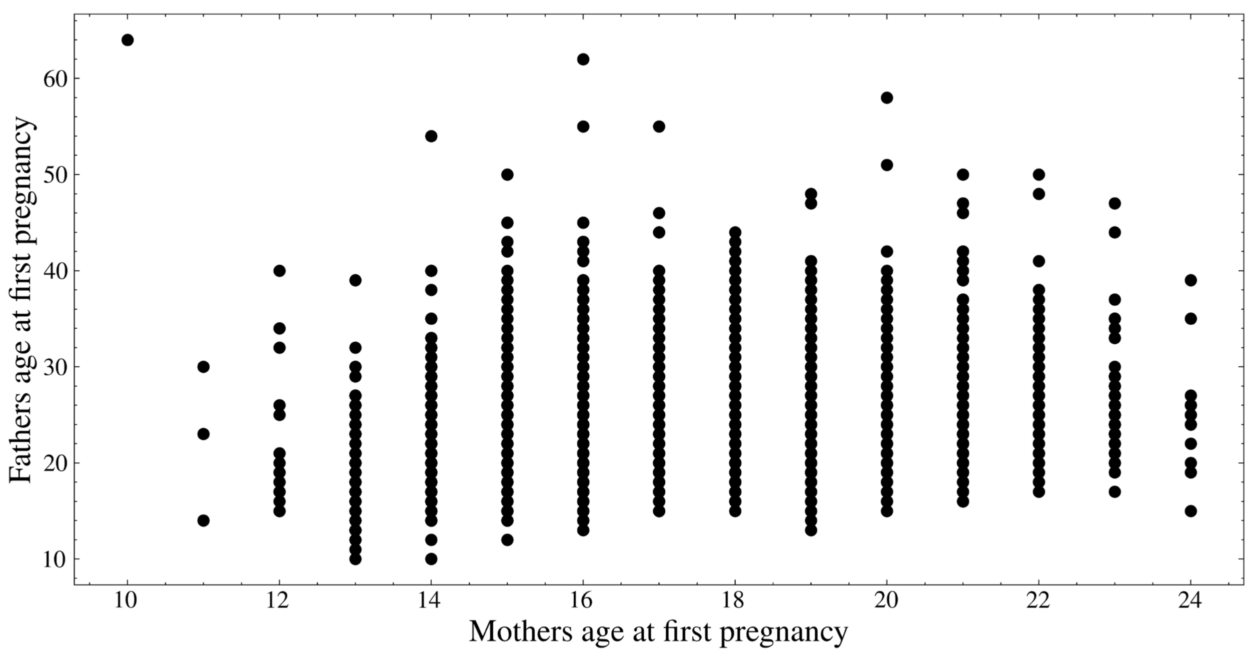


Figure A.5 Relationship between father's and mother's age of first pregnancy by teenage pregnancy

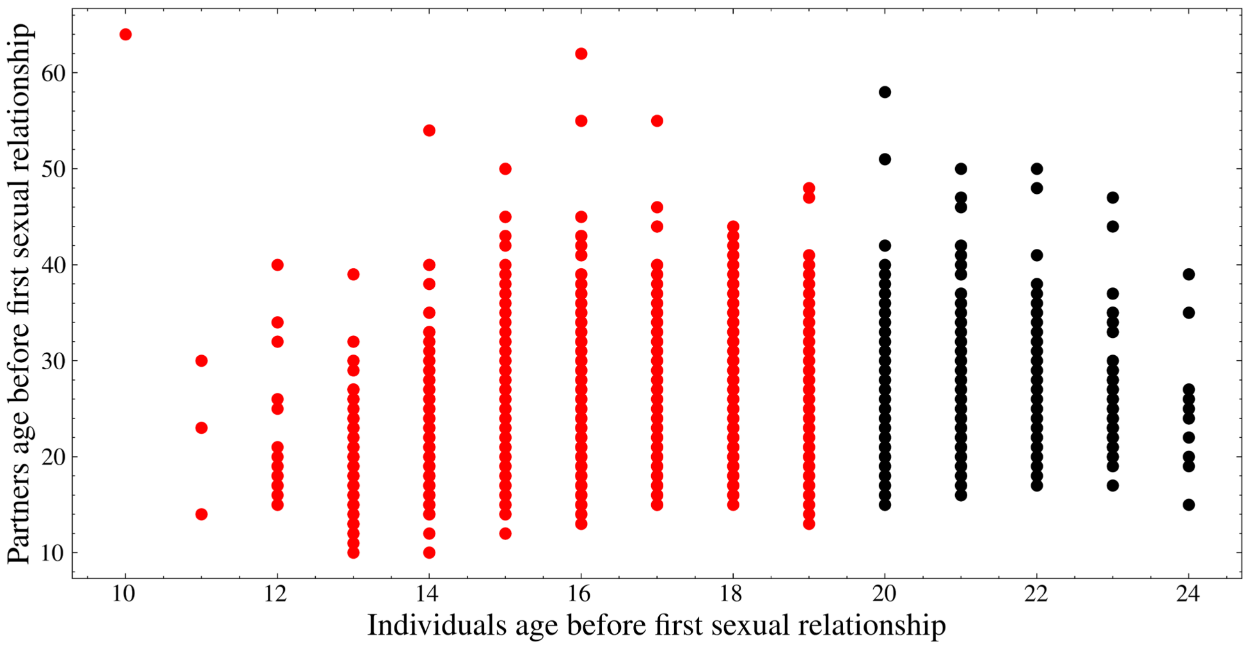
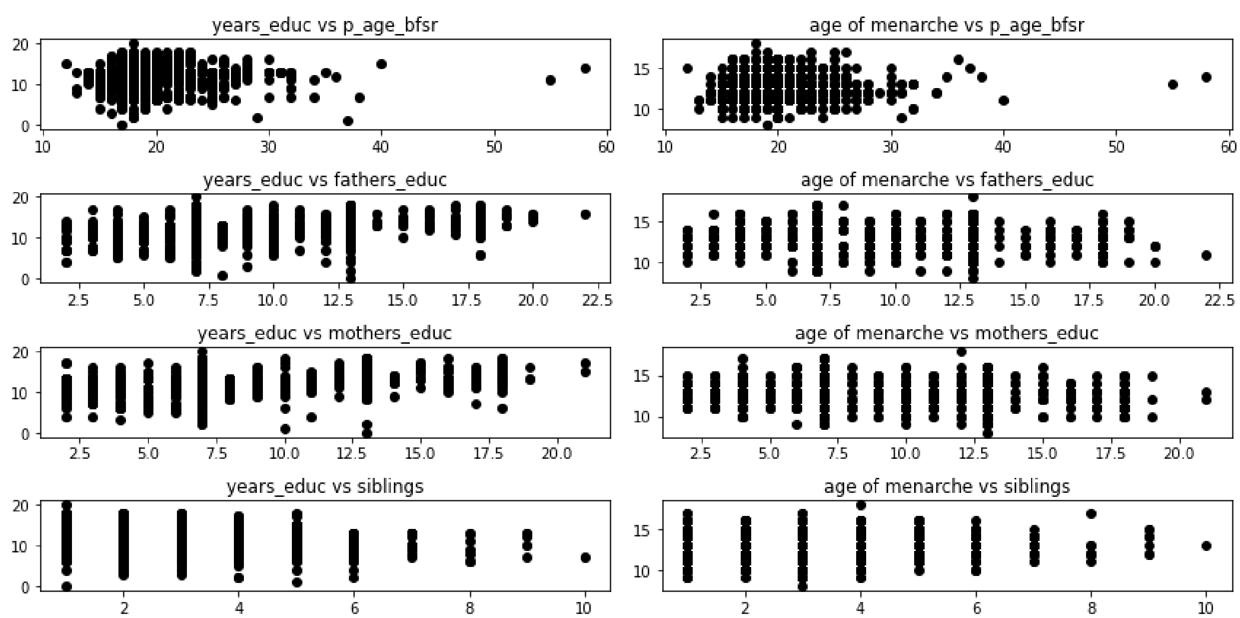


Figure A.6. Scatter plots between the dependent and instrument variables vs the covariates



8.2 Tables

Table A.1. Univariate descriptive statistics for secondary model

	Mean	Median	Std. D.	Min	Max	p25	p75	Skew	Kurt
years_educ	12.19	13	2.560	0	20	11	13	-.99	5.11
teen_pregnancy	.78	1	0.420	0	1	1	1	-1.34	2.81
menarche	12.69	13	1.440	8	18	12	14	.25	3.13
p_age_bfsr	19.78	19	3.760	12	58	18	21	2.76	20.59
contraceptive	.35	0	0.480	0	1	0	1	.61	1.37
highlands_region	.41	0	0.490	0	1	0	1	.36	1.13
coastal_region	.35	0	0.480	0	1	0	1	.63	1.4
amazon_region	.23	0	0.420	0	1	0	0	1.27	2.61
afro_ecuadorian	.05	0	0.220	0	1	0	0	4.2	18.61
mixed	.77	1	0.420	0	1	1	1	-1.25	2.57
white	.01	0	0.080	0	1	0	0	12.82	165.43
urban	.6	1	0.490	0	1	0	1	-.4	1.16
fathers_educ	9.13	7	3.760	2	22	7	13	.71	2.92
mothers_educ	8.86	7	3.790	2	21	7	13	.53	2.68
siblings	3	3	1.570	1	10	2	4	1.05	4.35

Table A.2. Univariate descriptive statistics for teenage mothers for secondary model

	Mean	Med.	S.D.	Min	Max	p25	p75	Skew	Kurt
years_educ	11.91	13	2.510	0	20	11	13	-1.06	5.29
menarche	12.63	12	1.410	8	18	12	14	.25	3.26
p_age_bfsr	19.49	19	3.630	13	55	17	21	2.38	15.5
contraceptive	.33	0	0.470	0	1	0	1	.72	1.52
fathers_educ	9.09	7	3.680	2	20	7	13	.72	2.93
mothers_educ	8.87	7	3.800	2	21	7	13	.52	2.69
siblings	3.04	3	1.630	1	10	2	4	1.02	4.13

Table A.3. Univariate descriptive statistics for non-teenage mothers for secondary model

	Mean	Med.	S.D.	Min	Max	p25	p75	Skew	Kurt
years_educ	13.15	13	2.500	4	18	13	14	-1.04	5.25
menarche	12.9	13	1.550	9	17	12	14	.17	2.74
p_age_bfsr	20.82	20	4.000	12	58	18	22	3.92	32.89
contraceptive	.44	0	0.500	0	1	0	1	.24	1.06
fathers_educ	9.26	7	4.020	2	22	7	13	.66	2.83
mothers_educ	8.82	7	3.730	2	18	7	13	.53	2.65
siblings	2.86	3	1.330	1	9	2	3	1.03	4.94

Table A.4. Correlations between dependent, endogenous and instrument variables with continuous covariates in secondary model

Variables	(1)	(2)	(3)
(1) years_educ	1.000		
(2) teen_pregnancy	-0.200	1.000	
(3) menarche	0.044	-0.075	1.000
(4) p_age_bfsr	0.005	-0.147	0.010
(5) fathers_educ	0.267	-0.019	-0.065
(6) mothers_educ	0.292	0.005	-0.086
(7) siblings	-0.207	0.047	0.079

Table A.5. Hyperparameter results for menarche and covariates for secondary model.

Hyperparameter Tuning Results							
Random Forest		**Bagging**		**SV Regression**		**Neural Network**	
n_estimators	100	n_estimators	200	C	0.5	hl_size	(7,7)
max_depth	5	max_depth	1	gamma	2^(-7)	alpha	10
min_s_split	1	min_s_split	1	epsilon	0.1		
min_s_leaf	2	min_s_leaf	10				

Table A.6. Hyperparameter results for teen pregnancy and covariates for secondary model.

Hyperparameter Tuning Results							
Random Forest		**Bagging**		**SV Regression**		**Neural Network**	
n_estimators	50	n_estimators	50	C	0.03125	hl_size	(3,3)
max_depth	5	max_depth	5	gamma	2^(-11)	alpha	0.1
min_s_split	5	min_s_split	5				
min_s_leaf	1	min_s_leaf	1				

Table A.7. Hyperparameter results for schooling years and covariates for secondary model.

Hyperparameter Tuning Results							
Random Forest		**Bagging**		**SV Regression**		**Neural Network**	
n_estimators	30	n_estimators	30	C	8	hl_size	(5,5)
max_depth	5	max_depth	1	gamma	2^(-7)	alpha	10
min_s_split	2	min_s_split	1	epsilon	0.1		
min_s_leaf	2	min_s_leaf	1				

Table A.8. Double ML Results for the secondary model

	OLS	**RF**	**BAG**	**SVR**	**NN**
coef	-6.333	-6.478	-6.541	-3.259	-6.586
std. error	2.706	3.542	3.736	1.812	3.327
t	-2.340	-1.829	-1.751	-1.799	-1.98
P>|t|	0.019	0.067	0.080	0.072	0.048
2.5 %	-11.637	-13.421	-13.863	-6.810	-13.106
97.5 %	-1.029	0.465	0.781	0.291	-0.066

Note: Standard-Errors are not robust to heteroscedasticity