THÈSE

En vue de l'obtention du

DOCTORAT DE L'UNIVERSITÉ DE TOULOUSE

Délivré par l'Université Toulouse 1 Capitole

Présentée et soutenue par

Jose MUNOZ ALVARADO

Le 25 août 2023

Essais sur l'économétrie appliquée et l'économie du travail

Ecole doctorale : **TSE - Toulouse Sciences Economiques**

Spécialité : **Sciences Economiques - Toulouse**

Unité de recherche :
TSE-R - Toulouse School of Economics - Recherche

Thèse dirigée par
Thierry MAGNAC et François POINAS

Jury

Mme Senay SOKULLU, Rapporteure
M. Thierry MAGNAC, Directeur de thèse
M. François POINAS, Co-directeur de thèse
M. Olivier DONNI, Président

Essays on Applied Econometrics and Labor Economics

Doctoral Dissertation

by

José Alfonso Muñoz Alvarado

August 25, 2023

Toulouse School of Economics

Université Toulouse Capitole 1

Acknowledgements

First and foremost, I want to thank my advisor, François Poinas, for his guidance throughout the years, and his advice, support, and devotion to assisting me with my Ph.D. There are no words to convey how much I learned and benefited from our several conversations. I also like to thank my other advisor, Thierry Magnac, for his many suggestions over the years, as well as for sharing and teaching me his exceptional insights. Along with François and Thierry, I also want to thank Marion Goussé, Olivier Donni, and Senay Sokullu for being members of the jury.

I am particularly grateful to Olivier de Groote for his insightful chats and assistance during my Ph.D. and job market. I am also grateful to numerous professors and staff in Toulouse who have assisted me in improving my work throughout the years. I thank Matteo Bobba, Nour Meddahi, Pepita Miquel, Louise Strachan, Vishal Kamat, and the members of the econometrics and empirical economics group.

I have also got the opportunity to share my work with others outside of TSE. I am grateful to Bram de Rock for hosting me at *Université Libre de Bruxelles* for three months, where I was able to join and meet outstanding researchers, which significantly improved my job market paper. I was also able to share my work with researchers at the *Universidad de Costa Rica*, for which I will be eternally grateful. Thanks, Yanira Xirinach, Juan Robalino, and Valeria Lentini provided me with valuable comments and encouragement. I would also want to express my gratitude to the *Instituto Nacional de Estadìsticas y Censos* and the *Centre d'études et de recherches sur les qualifications* for providing me the data that I needed to write my articles.

I had the honor of sharing my Ph.D. experience with Anaïs Fabre, Gökçe Gökkoca, Hippolyte Boucher, Léa Bignon, Lisa Botbol, Max Lesellier, Moritz Loewenfeld, Peter Neis, Stephan Waizmann, Tim Ederer, and Tuuli Vanhapelto, all of whom I love with all my heart. Thank you! I also thank Christophe Bruneel-Zupanc, Diego Dabed, Dhruv Jain, Filippo Maria D'Arcangelo, Guillem Focault, Gyung-Mo Kim, Ilaria Natali, Juan Pal, Kevin Remmy, Miguel Zerecero, Oscar Jara, Paloma Carrillo, Paul-Henri Moisson, Rossi Abi-Rafeh, Tanja Linta, Umutcan Salman, and many more fellow students from Toulouse for making this such a great time, and for help along the way.

I also thank many friends who I had the opportunity to meet and create friendships with all these years. Thank you Alea Muñoz, Anneliese Krautkraemer, August Aubach, Corbinian Schlosser, Daniel Ostalé, Fernanda Hernández, Mariana Odio, Michelle Amén, Paula Navarro, Paulina Rebolledo, Saí Bravo, and Valérie Furio. My chicos José Carlos Ortega and Simon Arthur. And my Costa Rican friends in Toulouse: Erick Mora, Esteban Muñoz, Esteban Vargas, Felipe Ramírez, Kerry Loaiza, María Evans, Sebastián Peña, and Viviana Meoño.

I would also like to thank my best friends María del Mar Vargas and Federico Ulloa, my aunts Seidy, Mery, Carmen, Roxana, and the rest of my family.

Last, this thesis is dedicated to people that gave me their unconditional support all the way: my mother Marlene, my siblings María Daniela and Juan Enrique, and Louis Vasseur, who in the last years has been my rock and motivation to give my best. I love you.

Overview

This doctoral thesis consists of three chapters that address questions related to the labor market and individual decision-making.

Chapter 1 focuses on the decision-making process of teens and parents in the allocation of time and income within households. By using data from the Costa Rican *Encuesta Nacional de Hogares* and the conditional cash transfer program *Avancemos*, the chapter identifies gender differences in household responses to the transfer using a marginal treatment effect framework and a collective household model. The results show that sons bargain cooperatively with their parents while daughters do not, indicating that sons have a higher opportunity cost of attending school than daughters. This has important implications for public policy targeting teens, as the gender disparity must be accounted for to be effective.

Chapter 2 assesses the impact of Costa Rica's Responsible Paternity Law on household decision-making. By using the law as a natural experiment and a fuzzy differences-in-differences setting, the chapter finds that the law had a negative impact on male labor participation as well as female and male weekly labor supply. Using a collective household model with matching, the chapter argues that the law strengthens women's bargaining power in household decision-making, leading to a couple selection effect and an intra-household allocation effect. The findings demonstrate how child-related laws can help better understand the household formation and decision-making.

The common reasoning behind the first two chapters of this thesis is the use of *collective household models*. The idea is that households collective bargaining how they allocate time and income across their members. Introducing multiple decision-makers in the household provides different results on how public policies affect society. Both first two chapters provide results on this, quantifying the effects of public policies on male and female labor outcomes.

The third chapter is coauthor with François Poinas. We propose a dynamic human capital accumulation model with two methodological improvements to examine the robustness of empirical data. We incorporate time-varying unobserved heterogeneity and endogenous attrition. Preliminary results show that attrition bias has no significant effect on estimated parameters. But it does affect sample composition over longitudinal panels and simulations by over-representing more-educated individuals with higher wages at older ages. The findings also show time-varying components in unobserved heterogeneity over time, in particular by accumulating schooling. It asses the importance of educational system tracking in skill accumulation and decision-making processes. We shed light on the diverse effects of accumulated schooling in skill acquisition and correcting for attrition in simulations that can be used for policy recommendations.

Contents

Chapter 1

Power to the teens? A model of parents' and teens' collective labor supply

Abstract

Teens make life-changing decisions while constrained by the needs and resources of the households they grow up in. Household behavior models frequently delegate decision-making to the teen or their parents, ignoring joint decision-making in the household. I show that teens and parents allocate time and income jointly by using data from the Costa Rican *Encuesta Nacional de Hogares* from 2011 to 2019 and a conditional cash transfer program. First, I present gender differences in household responses to the transfer using a marginal treatment effect framework. Second, I explain how the gender gap from the results is due to the bargaining process between parents and teens. I propose a collective household model and show that sons bargain cooperatively with their parents while daughters do not. This result implies that sons have a higher opportunity cost of attending school than daughters. Public policy targeting teens must account for this gender disparity to be effective.

1.1 Introduction

Teens make decisions that have lifelong implications for their human capital accumulation, potential income, and welfare (Cunha et al., 2006). These decisions are made along with their household's needs, such as housework or earning a wage, and may result in them dropping out of school. It raises concern since school dropouts earn less and have fewer work possibilities for the rest of their lives (De Hoyos et al., 2016). To avoid this, governments use subsidies such as conditional cash transfers (CCT), which give cash incentives to households as long as the teens continue to attend school. In addition to their benefits on educational outcomes in primary and secondary schools (García and Saavedra, 2017), CCTs have significant effects on households.

CCTs' have a direct impact on households by relaxing their budget constraints. Todd and Wolpin (2006) provide evidence that the Mexican CCT *Oportunidades* influences parental fertility decisions via this effect. Their unitary model assumes that households are single decision-making agents. However, this model does not account for the effect of CCTs on intrahousehold bargaining. To quantify this effect, a collective model of household decision-making in which multiple decision-makers allocate income, time, and consumption (Chiappori, 1992) is needed. De Rock et al. (2022) show how *Oportunidades* changes household consumption patterns, including those of young children, using such a model. This result is partly due to the mother's greater bargaining power and increased sharing with her children. Although teens have been included in collective models of consumption, the literature has ignored the possibility that they allocate time cooperatively with their parents, thereby overlooking the teens' preferences and opportunity costs.

In this paper, I present evidence that teens allocate time by bargaining cooperatively with their parents. Using the Costa Rican *Encuesta Nacional de Hogares* (ENAHO) from 2011 to 2019, I provide two sets of results. First, I present reduced-form evidence that a CCT program designed to keep teens in high school changes household time outcomes. Using a marginal treatment effect approach that accounts for the CCT program's endogeneity, I show that teens' schooling and parents' labor outcomes differ by gender for households that are indifferent between applying or not applying to the transfer program. If the teen is a daughter, her parents enjoy more leisure by decreasing their labor outcomes because of the program. Sons increase their school attendance, but only their mothers enjoy more leisure time, not their fathers. These gender differences in household responses to the CCT may be related to teens' bargaining power in the household. The second set of results investigates whether the teen bargains collectively with her parents about time allocation. I present a collective household model where teens and parents bargain to distribute income and time. Among other empirical results, I find that daughter does not cooperatively bargain with her parents. The son does, however, and is considered a decision-maker. I provide estimates of the bargaining function through which households with a teen son allocate resources. To the best of my knowledge, this is one of the first papers to consider the teen decision-maker in a time-use framework.

For my analysis, I evaluate the Costa Rican conditional cash transfer *Avancemos* on household decision-making. This CCT was established in 2006 and provides cash transfers to households as long as the teen remains in high school. The data I use comes from the Costa Rican ENAHO, a yearly household survey that collects individual data on a representative sample of households. It includes demographic, domestic, and labor market variables, and whether the household receives *Avancemos*. For the reduced-form results, I estimate a marginal treatment effect as presented by Heckman and Vytlacil (2005). As an instrument, I use the percentage of households that have *Avancemos* in the neighborhood. This variable is a proxy for treatment take-up spillover effects, accounting for the stigma, or willingness, associated with receiving social transfers and does not affect household outcomes. The instrument can be considered exogenous because the peer effect is on treatment take-up rather than on outcomes.

To rationalize gender differences in household responses to *Avancemos*, I present a collective household model in which households make decisions about teen high school participation, the labor supply of the parents, and domestic public good production. This model is built on Blundell et al. (2007) and Cherchye et al. (2012). I use the model predicted restrictions to find whether the teen is a decision-maker. The primary goal of the test is to compare how changes in wages and nonlabor income impact the outcomes of the parents and teens. In a unitary model with a single agent, transfers and earned wages have the same income effect on the household because they relax the budget constraint. The collective model with two or more decision-makers has an additional effect generated by transfers and wages because of the bargaining process. Aside from the income effect, a decision-maker who brings extra income to the household gains more gains bargaining power. My results show that although daughters are more effective at housework than sons, they do not have a bargaining position in the household. Because sons do negotiate with their parents, I can recover their bargaining function. The sons' share of income rises with additional transfers and their wages. However, they only receive a portion of their parents' wages, which are the primary source of income for the household. Last, as a result of the bargaining, the sons' income share decreases if they attend school but do not work.

These findings are useful in policy design. First, the opportunity cost of attending school for sons includes the bargaining effect, whereas it does not for daughters. Furthermore, daughters complement their education with more domestic work, which decreases their opportunity cost. Policies targeting teens must account for this gender difference to be effective. In addition, these effects may result in sons specializing in labor and daughters in domestic work. This specialization can have longer term effects on lifetime earnings and human capital accumulation as shown by women having lower labor participation even with higher education than men. As a result, public policies aimed at closing gender gaps should take gender roles into account from an early age in households.

The paper is structured as follows: In the next section, I discuss the literature. Section three gives *Avancemos*' background. The fourth section presents the data and empirical evidence on the impact of *Avancemos* on household time allocation decisions. I continue with the theoretical model in section five and empirical specifications of the unitary and collective models in section six. Results are shown in the seventh section. Section eight concludes.

1.2 Related literature

My paper adds first and foremost to the collective household model literature. Chiappori (1992) presents a study of household consumption and labor supply as a collective decision between the parents of a household. Chiappori and Mazzocco (2017) provide a comprehensive review of the literature. The literature concentrated on parents as decision-makers until the study by Dauphin et al. (2011). Their results provide compelling evidence that teens who are sixteen years old or older cooperatively bargain with their parents about consumption allocation. This result, together with the method presented by Dunbar et al. (2013), extended the literature by incorporating children as decision-makers. These analyses, however, only include consumption decisions and not labor supply. This is explained by the difficulty of observing children's wages and an exogenous variation in the nonlabor income of teens to identify their share of income as decision-makers. I contribute to the literature by developing a collective model and conducting empirical estimations with a teen as a decision-maker in a time-use framework using a rich dataset from Costa Rica and the CCT *Avancemos*.

My paper also builds on the literature on teen education decision-making. In a household setting, parents make decisions on behalf of the teen. Del Boca et al. (2014), for example, study parents' labor supply and the cognitive development process of children. They find that cash transfers to households with children have small impacts on child quality production because a significant fraction is spent on other household consumption and the leisure of the parents. Few articles investigate teens' schooling decisions in a collective model. Reggio (2011) and

Keshavarz Haddad (2017), for example, adopt a collective model where mothers decide their children's schooling or work decisions. However, Lundberg et al. (2009) argue that teens become capable of productive and independent work while continuing to depend on their parents. Ashraf et al. (2020) provide evidence that teens can negotiate their education. My paper is the first to use a collective bargaining approach in which teens make time-allocation decisions alongside their parents.

The paper adds to the literature on the effects of transfers in households. For example, Attanasio and Lechene (2014) show how the mother, the recipient of the transfer in *Oportunidades*, gains more control over household resources, affecting household consumption and labor market behavior. Dunbar et al. (2013) propose an estimation that allows the determination of the share of household resources for children. A large body of literature follows this approach, including Brown et al. (2018), Tommasi (2019), Calvi (2020) and Calvi et al. (2021). They extend the analysis by developing a measure of intrahousehold poverty. Furthermore, Sokullu and Valente (2022) analyze the effect of *Oportunidades* and suggest a novel identification strategy for recovering the household share of consumption, including children. In a similar topic, but without employing a collective model, Dubois and Rubio-Codina (2012) shows how CCTs make households reallocate care time for younger children between the mother and eldest daughter. I contribute to the literature by focusing on a transfer's effect on intrahousehold bargaining over the allocation of teens' time.

1.3 Institutional background

By 2012, Costa Rica had a secondary school dropout rate of 10%, higher than its primary school rate of 2.5% (Mata and Hernández, 2015). An explanation for this difference is higher opportunity costs of sending children to school rather than working in households with tighter budgets. In 2006, the Costa Rican government implemented *Avancemos*, a conditional cash transfer program to retain teens from low-income households in secondary education (Muñoz-Alvarado, 2016). Every public high school offers the possibility of applying to it. Once a household applies, *Avancemos* is assigned according to the Social Information Sheet (FIS) of the Mixed Institute for Social Aid (IMAS) (Mata and Hernández, 2015). The FIS collects household information and assigns a poverty score based on the wealth and demographics of the household. The household can apply for *Avancemos* if the teen is between the ages of 12 and 25 years old and is registered in high school[1] (Instituto Mixto de Ayuda Social de Costa Rica, 2022). Households

[1] Costa Rican high school system consists of 5 years, where the usual age for the student is from 13 to 17 years old. However, there are night high schools for older students who work during the day.

that receive *Avancemos* must send the teen to at least 80% of monthly classes. The amount of *Avancemos* initially varied across the five years of high school, increasing each year. Since 2015, the amount has been fixed at 42 US dollars[2] per month for the first three years of high school and 65 US dollars per month for the last two years[3] (Hernández Romero, 2016). In comparison, the monthly cost of a basic food basket in Costa Rica in July 2015 was 84 US dollars (Instituto Ncional de Estadísticas y Censos, 2015).

Some papers have investigated the effects of *Avancemos*. Meza-Cordero (2011) shows that it increases beneficiaries' education by half a year, two-thirds of a year for men. Mata and Hernández (2015) find an increase between 10 and 16% in high school retention due to the program. *Avancemos* also decreased child labor by approximately two percent between 2005 and 2006, with a greater effect on rural areas for male students (Lang Clachar et al., 2015). There have been no studies on the impact of *Avancemos* on parents' outcomes. For other children in the household, Muñoz-Alvarado (2016) finds no effect on the likelihood of beneficiary siblings' attending school. Common across these results are gender differences, with male students benefiting more than female students. One explanation is the difference in the opportunity cost of schooling between daughters and sons. According to Jiménez-Fontana (2015), women specialize in domestic work while men focus on market work in Costa Rica. This specialization may go hand in hand with schooling and domestic work being more complementary, lowering the opportunity cost for daughters to attend school compared with sons. My collective model includes this potential mechanism by modeling schooling, work in the labor market, and domestic time.

1.4 Data

I use the ENAHO repeated cross-section data from 2011 to 2019 (Instituto Ncional de Estadísticas y Censos, 2022). It consists of a representative sample of Costa Rican households, such as nuclear families, single parents without children, extended families, and so on. It collects information on the demographic, labor market, and domestic variables of each household member. It also includes information about the household, such as transfers and who receives *Avancemos*. The survey is ideal for my analysis for two reasons. First, the survey collects labor variables regardless of whether the person works in the formal or informal sector. It accomplishes this by assuring respondents that their information will not be shared with tax authorities, as well as by structuring the labor-related questions so that they do not directly ask about the in-

[2]The exchange rate used was the average of the selling exchange rate indicated by the Central Bank of Costa Rica for the first 8 months of 2015: 540.77 colones per US dollar

[3]Technical high schools in Costa Rica have an extra year and the amount of *Avancemos* is the same as the year before.

dividual's sector of employment. This characteristic allows me to observe teens' labor variables and wages, as they are more likely to be in the informal sector due to their low-skill level. Second, the ENAHO collects the number of hours people dedicate to housework (cleaning, chores, caring for younger children and the elderly).

The ENAHO sample consists of 100,989 households, with an average of 11,000 households each year. I select nuclear families (mother, father, and children) whose parents are aged 30 to 64, whose oldest child is between ages 15 and 20 years old, and with any other children 14 years old or younger. The main reason for this selection is to have only one child who is legally allowed to work[4] and no other adult in the household who can provide extra income. I also choose households in which the father works and eliminate observations with extreme values on wages, labor hours, and total household income[5].

Table 1.1 shows descriptive statistics of my sample. Most families have more than one child, are not poor[6], and live in a rural area outside the Central Valley or in an urban area in the Central Valley.

Table 1.1: Households' descriptive statistics

Variable	**Obs**	**Mean**	**Std. Dev.**	**Min**	**Max**
Total household income	2,447	163.67	77.62	30.31	480.61
Poor	2,447	0.20		0	1
Not poor	2,447	0.80		0	1
Number of children					
One	2,447	0.30		0	1
Two	2,447	0.42		0	1
Three or more	2,447	0.28		0	1
Geographic urban area					
Outside Central Valley, rural area	2,447	0.35		0	1
Central Valley, rural area	2,447	0.21		0	1
Outside Central Valley, urban area	2,447	0.17		0	1
Central Valley, urban area	2,447	0.27		0	1

Table 1.2 shows descriptive statistics of parents. Most of the parents have a school degree. There is a significant difference between fathers and mothers in the labor market and domestic hours. Mothers work an average of 40 hours per week in the home, while fathers work an average of 3.5 hours per week. However, fathers work an average of 55 hours per week in the

[4] 15 years old is the legal age to start working.

[5] All monetary variables are corrected for inflation with the base year being 2019. Additionally, I normalized the units to refer to one thousand colones, corresponding to 1.58 US dollars.

[6] Define in the ENAHO under a poverty line that assesses the household's ability to meet its needs through consumption.

labor market, while 34% of mothers have a job and work an average of 38 hours per week.

Table 1.2: Parents' descriptive statistics

	Father					Mother				
Variable	**Obs**	**Mean**	**Std. Dev.**	**Min**	**Max**	**Obs**	**Mean**	**Std. Dev.**	**Min**	**Max**
Age	2,447	44.57	7.13	30.00	64.00	2,447	40.61	6.10	30.00	62.00
Primary School diploma	2,447	0.66		0.00	1.00	2,447	0.66		0.00	1.00
High School diploma or more	2,447	0.17		0.00	1.00	2,447	0.17		0.00	1.00
Years of Schooling	2,447	6.96	3.15	0.00	17.00	2,447	7.14	3.14	0.00	17.00
Employed	2,447	1.00		1.00	1.00	2,447	0.31	0.46	0.00	1.00
Hourly wage*	2,447	1.78	0.83	0.17	4.62	770	1.66	0.90	0.10	4.70
Market hours*	2,447	49.86	11.93	8.00	81.00	770	38.37	15.52	6.00	80.00
Domestic participation	2,447	0.47		0.00	1.00	2,447	0.98		0.00	1.00
Domestic hours*	1,157	7.27	6.06	1.00	35.00	2,408	40.30	19.32	1.00	89.00

*: Conditional on participation.

Table 1.3 shows a clear gender difference in school attendance and domestic work supply for teens. Daughters are more likely than sons to help out around the house (86 percent vs. 62 percent), and if they do, they put in almost twice as many hours (11 hours per week vs 6 hours per week). On the other hand, sons are more likely to work and support the family financially and are less likely to attend high school.

Table 1.3: Teens' descriptive statistics

	Sons					Daughters				
Variable	**Obs**	**Mean**	**Std. Dev.**	**Min**	**Max**	**Obs**	**Mean**	**Std. Dev.**	**Min**	**Max**
Age	1,357	16.68	1.46	15.00	20.00	1,090	16.39	1.32	15.00	20.00
Age≥18	1,357	0.27		0.00	1.00	1,090	0.19		0.00	1.00
Decision										
Nothing	1,357	0.13		0.00	1.00	1,090	0.11		0.00	1.00
School	1,357	0.75		0.00	1.00	1,090	0.86		0.00	1.00
Paid work	1,357	0.13		0.00	1.00	1,090	0.02		0.00	1.00
Avancemos	1,357	0.28		0.00	1.00	1,090	0.35		0.00	1.00
Hourly wage*	170	1.26	0.59	0.20	3.92	26	1.17	0.39	0.35	1.99
Market hours*	170	44.85	14.58	8.00	78.00	26	40.62	16.48	13.00	80.00
Domestic participation	1,357	0.625		0	1	1,090	0.858		0	1
Domestic hours*	847	6.795	5.613	1	29	935	11.450	9.138	1	42

*: Conditional on participation.

1.4.1 Marginal Treatment Effect of *Avancemos*

To analyze the effect of *Avancemos*, I define the treated households as those who receive the transfer. In my sample, 70% of the 762 households that were treated had parents who only have a high school diploma, and 80% of those households have at least one other younger child. This suggests an endogenous treatment selection, where households with higher costs of sending their teen to school are those that receive *Avancemos*. I cannot control for treatment selection because I do not see which households apply for it or not in my data, only the households that benefit from it. To deal with this selection into treatment take-up, I estimate the marginal treatment effect developed by Heckman and Vytlacil (2005)[7]. From a policy point of view, this estimator enables us to know the benefits of *Avancemos* of households in the margin to apply or not by estimating its marginal costs and benefits. Knowing this makes it easier to adjust the subsidy to the policy's objectives and limitations.

Consider a household i that can receive the binary treatment of *Avancemos* ($D_i = 1$) or not ($D_i = 0$) and has a vector of covariates X_i. Its observed outcome Y_i is linked to the potential outcomes through the following equation:

$$Y_i = (1 - D_i)Y_{0i} + D_i Y_{1i}$$

The potential outcomes are specified as follows:

$$Y_{ji} = \mu_j(X_i) + U_{ji} \quad j = 0, 1,$$

where $\mu_j(\cdot)$ are unspecified functions and U_{ji} are random variables with mean zero conditional on covariates. The latent variable discrete choice model for selection into treatment is:

$$D_i^* = \mu_D(X_i, Z_i) - V_i$$
$$D_i = 1, \text{ if } D_i^* \geq 0, \quad D_i = 0, \text{ otherwise}$$

where Z_i is a vector of instruments and V_i is an i.i.d error denoting unobserved heterogeneity in the propensity for treatment. Because V_i enters the selection equation with a negative sign, it can be thought of as an unobserved distaste for being treated. Assuming F as the cumulative distribution function for V, the propensity score function is defined as $P(X_i, Z_i) = F_V(\mu_D(X_i, Z_i))$, and $U_{Di} = F_V(V_i)$ represents the quantiles of the distribution of the unobserved distaste for

[7] Blundell and Costa Dias (2009) and Cornelissen et al. (2016) provide a survey on the marginal treatment effect, its extensions, and applications.

treatment. Rewriting the latent variable model for selection into treatment as:

$$D_i = 1, \text{ if } P(X_i, Z_i) \geq U_{Di},$$

The marginal treatment effect on the household's outcomes of receiving *Avancemos* for households with observables $X_i = x$ and unobserved distaste of treatment $U_{Di} = p$ is:

$$\text{MTE}(x, p) = E[Y_{1i} - Y_{0i}|X = x, U_D = p] = \mu_1(x) - \mu_0(x) + E[U_1 - U_0|X = x, U_D = p].$$

The values p correspond to the values of the propensity score to receive *Avancemos*. I estimate the MTE where there is common support across the propensity score. The MTE can be interpreted intuitively as the average effect of treatment for households on the margin of indifference between applying or not applying to *Avancemos*. The indifference margin is set by the values of the propensity score. The propensity score values determine the indifference margin. The instrument's use allows for an exogenous variation, in which the change in treatment status captures the treatment effect for similar values of the propensity score. The similarities in the assumptions required to estimate the MTE and LATE estimators are explained by Cornelissen et al. (2016).

Instrument

Instruments that are related to distance are frequently used in literature. Carneiro et al. (2011), for example, use distance to the nearest high school to estimate the effect of educational returns on future wages. I am unable to create a distance-related variable with my data, but I can calculate the proportion of neighborhood households that receive *Avancemos* to get a proxy of the peer effect[8]. I use the entire ENAHO sample of 100,989 households to create this variable. I define 112 geographic blocks annually, each with an average of 100 households, and estimate the number of households in each household's block, excluding itself, that benefit from *Avancemos*. In my sample, households live in neighborhoods where on average 10% of households benefit from *Avancemos*.

For this instrument to be valid, it must fulfill two conditions: (i) The instrument Z_i is a random variable such that the propensity score $P(X_i, Z_i)$ is a nontrivial function of Z_i; and (ii) (U_{0i}, U_{1i}, V_i) are independent of Z_i, conditional on X_i. For condition (i), the instrument serves as a proxy for the household's positive or negative peer effects. If a middle-class household is the only one in the neighborhood to have the transfer, it might feel socially stigmatized as a

[8] DiPasquale and Glaeser (1999) uses a similar instrument in a home ownership framework.

result of getting it. A lower-income household can easily learn about and apply for *Avancemos*, though, if a lot of the homes in the neighborhood are receiving it. Table 1.4 shows that the instrument is relevant and has an impact on the uptake of the treatment.

Table 1.4: First stage: Propensity to receive *Avancemos*

	Daughters	Sons
Percentage of households with *Avancemos*	2.138***	1.449***
	(0.271)	(0.256)
Controls	Yes	Yes
Year and geographical effects	Yes	Yes
Observations	1,090	1,357
R^2	0.146	0.127

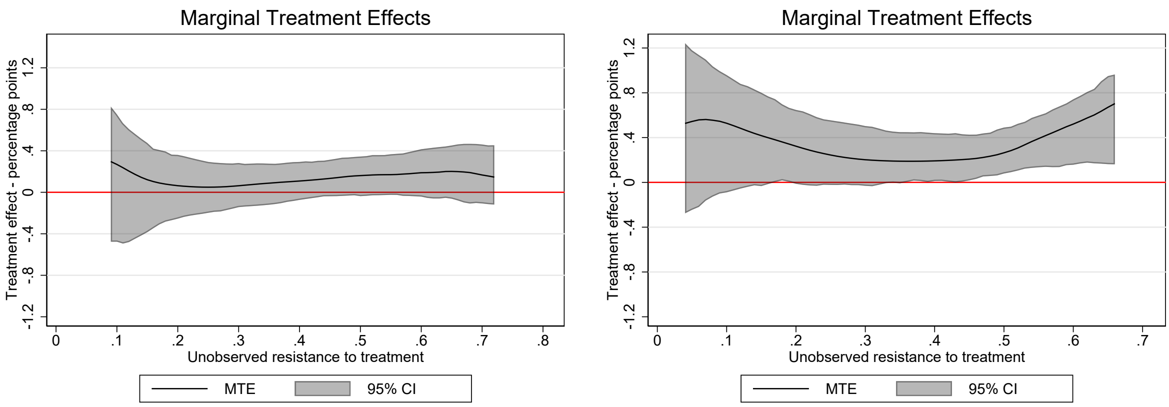
Standard errors between parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Control variables include the teens' age, parents' education, father's occupation, work industry, insurance, and the number of younger children in the household. The complete table is in Appendix 1.A.

The exogeneity of the instrument for condition (ii) is not obvious. Exogeneity should hold if the stable unit treatment value assumption (SUTVA) for the outcomes and the instrument's design, which takes spillovers in treatment uptake into account, are true. I control for a set of individual and household characteristics, including age, parental education, a marker for rural residence and region, and the father's job characteristics, such as occupation and industry, to strengthen the exogeneity. However, there can be concerns related to the endogeneity of where the household decides to leave. First, households are not located randomly. This endogeneity complements the idea of the instrument by taking into account the peer effects of the neighborhood in the treatment take-up. Secondly, I might be confusing the effect of *Avancemos* on labor outputs with the effect of infrastructure on labor outputs because neighborhoods with a lot of high schools can have a lot of other types of infrastructure. Even though this can bias my results, a more developed neighborhood has more private high schools that do not give the option to apply for *Avancemos*, hence moving in the same direction as my instrument.

Results

I estimate the MTE separately for the daughters and sons. To estimate the MTE for all quantiles, there should be enough units of treated and nontreated households for each value of the propensity score. However, Figure A1 shows that common support is not satisfied in my sample. I estimate the MTE in the daughters' sample up to the 69th percentile, and in the sons' sample up to the 65th percentile. My outcome variables are the teens' schooling decision, the fathers' labor supply, the mothers' employment status, and the domestic supply for the three members. Figure 1.1 shows the results for the unobservable part of the treatment effect for the teen's schooling decision[9]. The MTE has an upward-sloping shape for higher values of unobserved treatment resistance, indicating a pattern of reverse selection on gains. Whereas sons in households with a low resistance to *Avancemos* do not have a statistically significant effect on their attendance in high school, sons in households with higher resistance have a positive and significant effect. There is no significant effect on daughters. I show the estimation results of the observable part of the treatment effect in Appendix 1.A. The main differences in outcomes come from teens' age, mother's education, and father's occupation.

Figure 1.1: *Avancemos* on teen's schooling decision: MTE.



(a) Daughters.

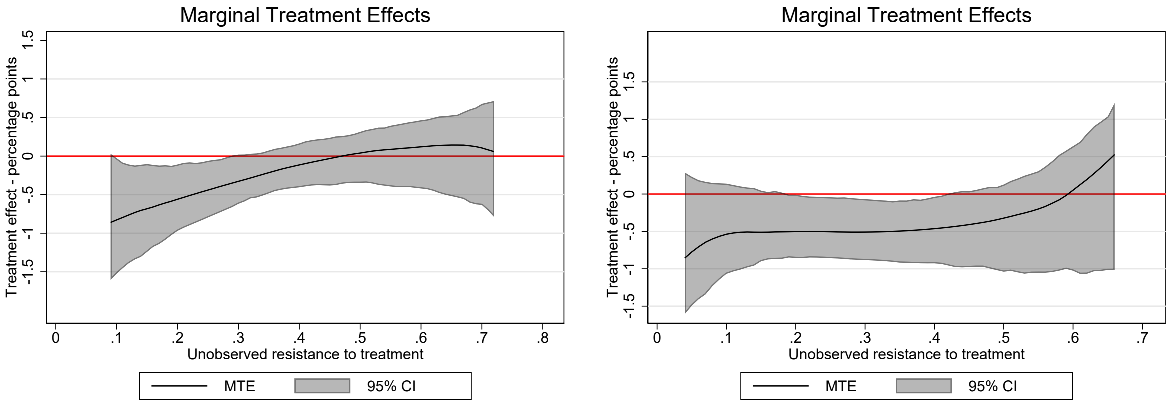
(b) Sons.

Figure 1.2 shows a negative and significant effect on the mother's employment status, regardless of whether she has a daughter or a son. Although the effects vary depending on the unobserved resistance, mothers in households with a higher resistance benefit most from *Avancemos*. Furthermore, the slopes are pretty different, indicating a negative selection for households with daughters but no selection on gains if the teen is a son (flat MTE). Figure 1.3 shows the results of the unobserved part of the MTE for the father's labor supply. For the fathers, there is no gender difference selection into gains, but fathers with a daughter benefit from *Avancemos* by decreasing their labor supply. Finally, none of the three household members experiences any significant changes in domestic supply. The graphs and the rest of the results are in Appendix

[9]Note: MTEs are calculated with the separate approach by Heckman and Vytlacil (2005) using a polynomial of degree 1 and a semiparametric local polynomial of degree 2. Standard errors are computed with a bootstrap procedure (250 replications).

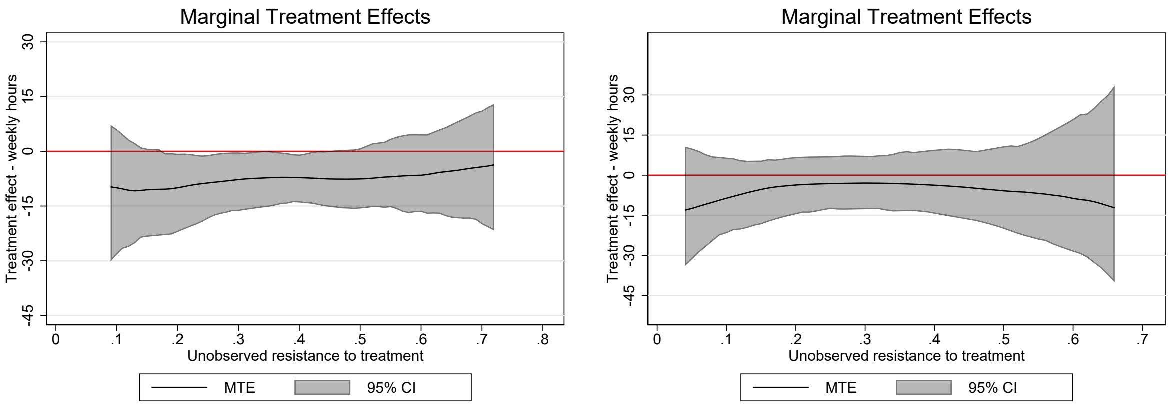
1.A.

Figure 1.2: *Avancemos* on mother's employment status: MTE.



(a) Daughters. (b) Sons.

Figure 1.3: *Avancemos* on father's labor supply: MTE.



(a) Daughters. (b) Sons.

To summarize, *Avancemos* reallocates household time. The daughter continues to attend high school, her mother is less likely to work, and her father reduces his labor supply. A standard income effect can explain these changes, in which *Avancemos*, as extra income, allows the parents to enjoy more leisure. However, it is not the same if the teen is a son. In that case, he is more likely to attend high school, and his mother is less likely to work in the labor market, but his father's labor supply remains unchanged. One explanation could be that the son bargains cooperatively with his parents and receives a share of the household's resources. If this is the case, the extra income from *Avancemos* is shared differently between parents and sons. To test whether the son is a decision-maker and to quantify his bargaining effect, I present a collective household model in the section below.

1.5 Theoretical model

The common models of household behavior are the unitary household model, which assumes that households behave as one single decision-maker, and the collective household model, which considers households as a group of individuals, each with their utility. In a collective household model, members allocate resources according to their bargaining power, which depends on their outside options as in a cooperative game.

The existence of cooperative bargaining in households is documented in the literature, but it excludes the possibility of teens bargaining with their parents about time allocation. Children are typically regarded as public goods in the utility of their parents. My data, however, show behavior that does not fit the traditional father and mother collective model. In this section, I present a collective model with parents and teens as decision-makers and a unitary model in a time-use framework. The models allow me to test the income pooling-hypothesis of the unitary model and the bargaining constraints obtained from the collective model.

The models include parents' labor supply, teens' schooling decisions, and the production of a domestic public good. Let $i = p, t$ denote the parents[10] and the teen. The household produces a public good through the production function $f^K(h^p, h^t)$, which takes domestic work (h^p, h^t) as input. This function is twice continuously differentiable in all its arguments, strictly increasing and strongly concave. The good created from this production function can be interpreted as the satisfaction of a clean and warm house. The household faces the same restrictions in the unitary and collective models. I present these restrictions first. For the parents, their time constraint[11] is between leisure, l^p; labor market work, m^p with the father always working, and domestic work, h^p:

$$l^p + m^p + h^p = 1.$$

The teen allocates her time between school $(s^t = 1)$, work $(s^t = 0)$ and domestic work h^t.

$$l^t + h^t + S\mathbf{1}\{s^t = 1\} + \overline{m}^t\mathbf{1}\{s^t = 0\} = 1$$

where S is the amount of time spent in school and $\overline{m}^t$ her working time. Each member consumes a Hicksian composite good C $(= C^p + C^t)$. The price of the consumption good is set to 1. The

[10]I combine parents as a single decision-maker for two reasons. First and foremost, my goal is the bargaining between parents and teens, not about disentangling fathers and mothers. Second, data constraints prevent me to consider a collective model with three decision-makers.

[11]Time is normalized to 1.

household acquires goods according to the following budget constraint:

$$C = w^p m^p + w^t \overline{m}^t \mathbf{1}\{s^t = 0\} + y + y_A \mathbf{1}\{Avancemos = 1\}\mathbf{1}\{s^t = 1\}$$

where w^p is the parents' hourly wage, w^t is the teen's hourly wage if she works, y is nonlabor income and y_A is *Avancemos*' transfer.

For the unitary model, the household behaves as a single unit according to a twice continuously differentiable, strictly monotonic, and strongly concave utility function $U^H(l^p, l^t, C, f(h^p, h^t))$. It maximizes the following program:

$$\max_{m^p, s^t, h^p, h^t, C} \quad U^H(l^p, l^t, C, f^K(h^p, h^t)) \tag{A17}$$

$$s.t. \begin{cases} C = w^p m^p + w^t \overline{m}^t \mathbf{1}\{s^t = 0\} + y + y_A \mathbf{1}\{Avancemos = 1\}\mathbf{1}\{s^t = 1\} & \text{(1.1a)} \\ 0 < m^p \leq 1, \quad s^t \in \{0,1\}, \quad m^t \in \{0, \overline{m}\} & \text{(1.1b)} \\ l^p + m^p + h^p = 1 & \text{(1.1c)} \\ l^t + h^t + \overline{m}^t \mathbf{1}\{s^t = 0\} + S\mathbf{1}\{s^t = 1\} = 1 & \text{(1.1d)} \end{cases}$$

Following Chiappori (1992), I assume in the collective model that each member has an egoistic, twice continuously differentiable, strictly monotonic, and strongly concave utility function $U^i(l^i, C, f^K(h^p, h^t))$. The members' bargaining function depends on the parents' labor income w^p, the teen's wage w^t, and the household's nonlabor income y. As is usual in the literature, I assume that the members choose Pareto efficient intrahousehold allocations [12]. The household's maximization program is:

$$\max_{m^p, h^p, C^p, m^t, h^t, C^t} \quad \lambda^p(w^p, w^t, y) U^p(l^p, C^p, f^K(h^p, h^t)) + \lambda^t(w^p, w^t, y) U^t(l^t, C^t, f^K(h^p, h^t)) \tag{2.2}$$

$$s.t. \begin{cases} \lambda^p(w^p, w^t, y) + \lambda^t(w^p, w^t, y) = 1 & \text{(1.2a)} \\ C^p + C^t = w^p m^p + w^t \overline{m}^t \mathbf{1}\{s^t = 0\} + y + y_A Pr(Avancemos = 1)\mathbf{1}\{s^t = 1\} & \text{(1.2b)} \\ 0 < m^p \leq 1, \quad s^t \in \{0,1\}, \quad m^t \in \{0, \overline{m}\} & \text{(1.2c)} \\ l^p + m^p + h^p = 1 & \text{(1.2d)} \\ l^t + h^t + \overline{m}^t \mathbf{1}\{s^t = 0\} + S\mathbf{1}\{s^t = 1\} = 1 & \text{(1.2e)} \end{cases}$$

where $\lambda^i(w^p, w^t, y)$ is the Pareto weight for the parents and the teen, which has a direct relationship to their bargaining power.

[12] A simple argument in favor of this assumption is that members are aware of each other's preferences and are unlikely to disregard Pareto-improving decisions as a result of their interaction. For more information about the validation of Pareto efficiency, see the surveys Vermeulen (2002) and Chiappori and Mazzocco (2017).

A two-stage model can be used to represent both the unitary and collective models. The first stage of the unitary model determines how much public good is produced. In the second stage, the household maximizes utility with residual income, conditional on the quantity of public goods produced. The first stage in the collective model is similar; the household agrees to produce a certain level of public good $\bar{f}^K$. However, conditional on the public good created and cooperative bargaining, parents and teens receive a share of the residual income. This distribution takes place at the Pareto frontier of household decision-making and is represented with the conditional sharing rule $\rho^i(w^p, w^t, y|\bar{f}^K),\ i = \{p, t\}$. In the second stage, each member decides freely on their level of leisure and consumption with their conditional share of income.

It is important to emphasize two points about these models. First, public good production is efficient in both cases, and the Bowen-Lindahl-Samuelson[13] condition holds (Blundell et al., 2005). Second, the difference in household behavior is reflected in both the teen's school attendance and the labor supply of the parents. Income is pooled in the unitary case, such that an increase in nonlabor income or in the teen's wage has the same effect on the parents' labor supply (if the teen only works). The extra income has a standard income effect on the teen's schooling decision. In the collective model, changes in a household's income include the bargaining effect. Because the teen's wage increases the teen's bargaining power, there is a difference in the parents' labor supply depending on whether extra income comes from nonlabor income or an increase in the teen's wage. There is also an effect even if the teen attends school and does not have a wage because her potential wage gives her bargaining power. The bargaining effect is reflected in the reservation wage that defines the teen's schooling and work decisions. Changes in her parents' wages have an impact on the reservation wage due to changes in the bargaining process, and thus her share of resources. The mathematical representation of these effects is in Appendix 1.B.1 for the unitary model and in Appendix 1.B.2 for the collective model.

1.5.1 Identification

The distribution of wages and nonlabor income are the primary sources of identification in the models. They are sufficient to identify all of the parameters in the unitary case. In the collective model, however, identifying the sharing rule is more difficult. I use two results from the literature to identify the sharing rule. First, Cherchye et al. (2012) provide the identification of the home production function within a collective model. The issue is that, in general, the variation in wages and nonlabor income also impacts domestic production. A solution is to use

[13]This condition states that at the optimal level of the public good, the sum of the parents and teen's marginal benefit from the public good is equal to its marginal cost.

"production shifters", variables that impact the home production function but not preferences. These variables enable estimating the substitution parameter in a CES function, fixing the level of public goods produced in the household. Once domestic production is fixed, the wage and nonlabor variation identify the bargaining function (Chiappori, 1992). The children's ages and house characteristics are frequently used as production shifters in literature. I am unable to use these variables because I condition my sample based on the teen and younger children's ages. I currently impose Cobb Douglass production function with a substitution parameter equal to one. It allows me to identify the model without using production shifters.

The second identification result is provided by Blundell et al. (2007) for collective models in the case in which one of the outputs is on the extensive margin, the teen's schooling decision in my case. To recover the sharing function from the estimation, it is necessary to make a double indifference assumption. This assumption states that parents and teens are indifferent about going to school across the schooling frontier. This assumption allows for the sharing rule to be continuous across the frontier, keeping the Pareto assumption across it. Finally, the conditional sharing rule is identified up to a constant. Because I do not observe individual private consumption or the output of the household production function, it is impossible to distinguish between household heterogeneity in outputs and sharing rules (Chiappori, 1992).

Both results provide nonparametric identification. Following the literature, I use a parametric specification to estimate the model.

1.6 Parametric specification

In this section, I describe the parametric model that allows me to estimate and test the unitary and collective models. Using the models' two-stage framework, I derive the equations I estimate and the constraints I test. I start with the estimation of the public good production, and then parental labor supply and teen schooling.

For the public good production, I assume it is produced with Cobb Douglas technology with a productivity parameter α:

$$f^K(h^p, h^t) = (h^p)^\alpha \ \ (h^t)^{1-\alpha}$$

and has a price equal to:

$$g^K(w^p, w^t) = \left(\frac{w^p}{\alpha}\right)^\alpha \ \left(\frac{w^t}{1-\alpha}\right)^{1-\alpha}.$$

If households produce effectively while minimizing costs, regardless of their behavior, the amount produced satisfies the Bowen-Lindahl-Samuelson condition of efficient public good provision.

Under this assumption and applying Shepard's lemma, the domestic supply Hicksian demands are:

$$\begin{aligned} h^p(w^p, w^t) &= \left(\frac{w^p}{\alpha}\right)^{\alpha-1} \left(\frac{w^t}{1-\alpha}\right)^{1-\alpha} \overline{f}^K \\ h^t(w^p, w^t) &= \left(\frac{w^p}{\alpha}\right)^{\alpha} \left(\frac{w^t}{1-\alpha}\right)^{-\alpha} \overline{f}^K. \end{aligned}$$

Taking the differences in log, I obtain the following equation:

$$\ln\left(\frac{h^p}{h^t}\right)(w^p, w^t) = \ln\left(\frac{w^t}{w^p}\right) + \ln\left(\frac{\alpha}{1-\alpha}\right) + u_h, \tag{1.3}$$

where u_h denotes a measurement error. This equation estimates the productivity parameter α.

1.6.1 Unitary model

For the unitary model, I assume a semilog indirect utility function for the household defined as:

$$v^H(w^p, y^*, \overline{f}^k) = \frac{\exp(\theta^y w^p)}{\theta^y}\left(\theta^p_w \ln w^p + \theta^y y^* + \theta^H_K \ln \overline{f}^K\right) - \frac{\theta^p_w}{\theta^y}\int_{-\infty}^{\theta^y w^p} \frac{\exp(t)}{t} dt, \tag{1.4}$$

where $y^* = y + w^t$ if the teen works or $y^* = y + y_A \mathbf{1}\{Avancemos\}$ if the teen goes to school. The Bowen-Lindahl-Samuelson condition in this model is:

$$g^K(w^p, w^t) = \frac{1}{\overline{f}^K}\frac{\theta^H_K}{\theta^y}$$

For the second-stage model, the Marshallian labor supply for the parents follows Roy's identity in equation (1.4):

$$M^p(w^p, w^t, y, \overline{f}^K) = \theta^p_w \ln w^p + \theta^y y^* + \theta^H_K \ln \overline{f}^K.$$

I replace $\overline{f}^K$ with the Bowen-Lindahl-Samuelson condition and obtain:

$$M^p(w^p, w^t, y) = \theta^p_w \ln w^p + \theta^y y^* + \theta^H_K\big[\alpha \ln w^p + (1-\alpha)\ln w^t - \alpha\ln\alpha - (1-\alpha)\ln(1-\alpha)\big]. \tag{1.5}$$

This equation is estimated using a switching regression model based on the teen's educational status:

$$\begin{aligned} m^p(w^p, w^t, y) &= A^*_p \ln w^p + A_t w^t + A_y y + \delta \ln w^t + \mathbf{X}_w\beta_w + u_w, \quad \text{if } s^t = 0 \\ m^p(w^p, w^t, y) &= a^*_p \ln w^p + a_t w^t + a_y y + \delta \ln w^t + \mathbf{X}_s\beta_s + u_s, \quad \text{if } s^t = 1 \end{aligned} \tag{1.6}$$

where $\mathbf{X}$ is a vector of covariates and u_w and u_s are measurement errors. For the teen's schooling decision, I use a latent index:

$$\begin{aligned} (s^t)^* &= b + b_t w^t + b_p \ln w^p + b_y^t y + \mathbf{X}_t \beta_t + u_t \\ s^t &= 1 \ \text{ if } \Phi((s^t)^* \geq 0), \quad s^t = 0 \ \text{ otherwise} \end{aligned} \tag{1.7}$$

u_t represents a measurement error. There are two key points to note in this equation. First, because it represents the schooling decision frontier, the public good is not included. Second, it is possible to recover the reservation wage when $s^t = 0$, where the frontier parameters are:

$$\gamma_p = -\frac{b_p}{b_t}, \quad \gamma_y = -\frac{b_y}{b_t}. \tag{1.8}$$

The novelty of my model is the introduction of home production in the estimation. It establishes a direct relationship between the reduced-form parameters and the Cobb-Douglas technology. Explicitly, in the estimation, $\delta = \theta_K^H \, (1-\alpha)$ and $A_p^* = A_p + \theta_K^H \, \alpha$, where A_p has a direct relationship to equation (1.5). Similarly, for a_p^*. Once the technology parameters are recovered, the remaining parameters are used to test the unitary model restrictions (Blundell et al., 2007):

$$\begin{aligned} A_t &= A_y \\ a_t &= 0 \end{aligned} \tag{U1}$$

$$\begin{aligned} (1+\gamma_y)(a_y - A_y) &= 0 \\ A_y \gamma_p &= (1+\gamma_y) a_p^* - A_p^* \end{aligned} \tag{U2}$$

The unitary model can be tested using these two sets of constraints. The null hypothesis for these restrictions is that the household acts as a single unit decision-maker. Under the null, restrictions U1 refer to the household income distribution and how it affects the labor supply of the parents. There is the same effect of an increase in the teen's wage and non-labor income on the parents' labor supply. If the teen does not work, there is no effect on her potential wage. Restrictions U2 refer to income and substitution effects caused by shifts in the labor supply of the teen's parents on her decision to enroll in school. Increases in non-labor income result in higher reservation wages, allowing the teen to continue their education for longer. The same holds for a raise in the parents' salary.

1.6.2 Collective model

The cooperative resource bargaining between parents and teenagers is the collective model's key characteristic. This negotiation establishes Pareto weights, which are converted into a unique resource share (ρ^i) for each household decision-maker. I assume that the parents' preference for market work and consumption has the following indirect utility belonging to the semilog labor supply equations (Stern, 1986):

$$v^p(w^p, \rho^p, \overline{f}^k) = \frac{\exp(\theta^p_\rho w^p)}{\theta^p_\rho}\left(\theta^p_w \ln w^p + \theta^p_\rho \rho^p + \theta^p_K \ln \overline{f}^K\right) - \frac{\theta^p_w}{\theta^p_\rho}\int_{-\infty}^{\theta^p_\rho w^p} \frac{\exp(t)}{t} dt. \tag{1.9}$$

The teen's indirect utility is:

$$v^t(w^t, \rho^t, \overline{f}^k) = \theta^t_\rho \rho^t + \theta^t_K \ln \overline{f}^K. \tag{1.10}$$

These utilities define the production of the public good and the conditional distribution of resources (Blundell et al., 2005). Under efficient production, the Bowen-Lindahl-Samuelson condition is:

$$g^K(w^p, w^t) = \frac{1}{\overline{f}^K}\left(\frac{\theta^p_K}{\theta^p_\rho} + \frac{\theta^t_K}{\theta^t_\rho}\right).$$

I assume the conditional bargaining function has the form

$$\rho(w^p, w^t, y|\bar{f}^K) = \psi_p \ln w^p + \psi_t w^t + \psi_y y, \tag{1.11}$$

where I use the level of teen earnings rather than the log because it allows me to nest the income pooling hypothesis with nonlabor income, the use of the log in the parents' wage is to connect it to the parents' labor supply. I assume for the second stage problem that $\rho(w^p, w^t, y|\bar{f}^K) = \rho^t$, which implies $\rho^p = y^* - \rho^t - g^K(w^p, w^p)\overline{f}^K$. Applying Roy's identity to the parents' indirect utility function gives their (conditional) Marshallian labor supply:

$$M^p(w^p, \rho^p, \overline{f}^K) = \theta^p_w \ln w^p + \theta^p_\rho \rho^p + \theta^p_K \ln \overline{f}^K$$

Substituting the sharing rule and including the Bowen-Lindahl-Samuelson condition gives:

$$M^p(w^p, w^t, y^*) = (\theta^p_w - \theta^p_\rho \psi_y + \theta^p_K \alpha)\ln w^p + (\theta^p_\rho(1-\psi_t))w^t + (\theta^p_\rho(1-\psi_y))y + \theta^p_K(1-\alpha)\ln w^t + U$$

where $U = \theta^p_K \ln\left(\frac{\theta^p_K}{\theta^p_\rho} + \frac{\theta^t_K}{\theta^t_\rho}\right) - \theta^p_\rho\left(\frac{\theta^p_K}{\theta^p_\rho} + \frac{\theta^t_k}{\theta^t_\rho}\right) + \theta^p_K(-\alpha\ln\alpha - (1-\alpha)\ln(1-\alpha))$. As in the case of the unitary model, I estimate this equation through a switching regression regime defined by the teen's schooling status:

$$\begin{aligned} m^p(w^p, w^t, y) &= A_p^* \ln w^p + A_t w^t + A_y y + \delta \ln w^t + \mathbf{X}_w \beta_w + u_w, \;\; \text{if } s^t = 0 \\ m^p(w^p, w^t, y) &= a_p^* \ln w^p + a_t w^t + a_y y^* + \delta \ln w^t + \mathbf{X}_s \beta_s + u_s, \;\; \text{if } s^t = 1 \end{aligned} \tag{1.12}$$

Finally, the teen's schooling decision is modeled as in the unitary case with equation (1.7). It is important to highlight that these reduced-form equations are the same as in the unitary model. However, depending on which model we accept or reject, the estimated parameters have different interpretations. Again, the novelty here is the introduction of home production. Using direct mapping to recover Cobb Douglas technology, the rest of the parameters are used to test the restrictions of the collective model, as in Blundell et al. (2007):

$$\frac{A_t - a_t}{A_y - a_y} = -\frac{1}{\gamma_y}, \quad \frac{A_p^* - a_p^*}{A_y - a_y} = \frac{\gamma_p}{\gamma_y} \tag{C1}$$

Restrictions C1 allow us to test the collective model. The household behaves as though there are two decision-makers, the parents and the teen, according to the collective model's null hypothesis. In contrast to the unitary model, households in a collective framework consider how non-labor income and wages may alter the household's budget constraint and the bargaining power of the decision-maker who brings them. Restrictions C1 are a simplified version of the restrictions in the parents' labor supply and the teen's schooling decision.

1.6.3 Stochastic specification

The estimation consists of four equations: the home production function, the teen's schooling participation, and the parents' switching labor supply. To allow for unobserved factors within households, I follow Bonnal et al. (1997) and assume that the four measurement errors (u_h, u_t, u_w, u_s) are generated by a common normally distributed random variable η such as:

$$u_{i,j} = \eta_i \chi_j + \varepsilon_{i,j}$$

where i is a household, j is one of the equations and $\varepsilon_{i,j}$ is an i.i.d. shock.

An important part of the paper is *Avancemos*, which is endogenous in this estimation for two different reasons. First, the households who benefit from it must send their teen to school because it is the condition of the transfer. Because of this, I only take into account the households that do not receive the transfer when I estimate the participation in schooling. Second, as *Avancemos* affects the amount of nonlabor income in the household, using this variable generates an omitted endogeneity bias due to the selection into treatment. I employ the control function method suggested by Blundell and Powell (2003) to correct this bias. I use the amount of *Avancemos* and household covariates to regress nonlabor income. I then use the residuals as a control variable in the parents' labor supply equations and the teen's choice of school. With the residuals serving as a covariate to correct the inconsistency of my estimates, much like a Heckman two-step correction, I treat endogeneity as an omitted variable problem in this way. Appendix 1.D contains the regression results in detail.

Last, I impute the teen's wages to estimate the model for those households in which the teen goes to school. For the imputation, I use a larger subsample from the ENAHO because there are not enough observations with wages in my sample, especially for daughters. It consists of individuals considered children in the household, aged 15 to 25, without a high school diploma. It differs from my sample in that it covers a variety of households, such as single parents and extended households. I estimate wages individually for sons and daughters. I explain this imputation in detail in Appendix 2.C.

I use maximum likelihood to estimate the unrestricted model, the unitary model, and the collective model. The likelihood contribution for household i is:

$$L_i = \left[\Pr((s^t)^* \geq 0) \times f(m_i^p | (s^t)^* \geq 0)\right]^{\mathbf{1}\{s_i^t=1\}} \times \left[\Pr((s^t)^* < 0) \times f(m_i^p | (s^t)^* < 0)\right]^{\mathbf{1}\{s_i^t=0\}} \times f(\ln(h_i^p / h_i^t))$$

1.7 Results

The main result is determining whether the teen is a decision-maker. For this, I use the likelihood ratio test for the unitary restrictions U1 and U2 and the collective restrictions C1. The null hypothesis for the unitary model is that the household acts as a single unit decision-maker. The null hypothesis for the collective model is that the household behaves with two decision-makers: the parents and the teen. The alternative hypothesis is not clear for both tests because it is difficult to tell from which channel the rejection of the null hypothesis comes. The estimation results are presented in Appendix 1.E. Starting with the son, the likelihood ratio statistic for the unitary model restrictions is 26.73 with a p-value of 0.00003; hence, I reject that a household with a son behaves as the unitary model predicts. For the collective model restrictions, the likelihood ratio statistic is 1.28 with a p-value of 0.53; hence, I do not reject that the son is a decision-maker. For the daughter[14], the likelihood ratio statistic for the unitary model is 21.98

[14] I estimate the domestic outputs for sons both jointly and separately with the other outcomes, and I obtain equivalent results. I assume that the same result would be valid for daughters. I do this because I impute most of the wages for the daughters, concentrating variation around the mean.

with a p-value of 0.002; hence, I reject that a household with a daughter behaves as the unitary model predicts. For the collective model, the likelihood ratio statistic is 6.93 with a p-value of 0.03; hence, I also reject the daughter as a decision-maker.

The difference between daughters and sons being decision-makers or not is reflected in their opportunity costs. Going to school suggests that the son has less negotiating power with his parents, which lowers his income share. The opportunity cost for daughters is at the household level and not on her own because the daughter lacks bargaining power and attends school as part of the decision made by the household. This opportunity cost is in the model via a reservation wage. Recovering the reservation wages for sons and daughters is possible with equation A4:

$$w^r_{son} = \kappa_{son} + \underset{(0.070)}{0.405}\ \ln w_p + \underset{(0.158)}{0.493}\ y, \quad \text{if teen is a son}$$

$$w^r_{daughter} = \kappa_{daughter} + \underset{(0.410)}{0.359}\ \ln w_p + \underset{(0.430)}{0.668}\ y, \quad \text{if teen is a daughter.}$$

The coefficients show that daughters' reservation wages increase more than sons' for additional non-labor income and the parents' wages, even though I cannot test the difference between them (because they were estimated using different samples).

My results show that daughters are not included in the decision-making process regarding their potential labor-force participation. Most daughters seem to be led to prioritizing their education over work. This could be because work opportunities are less appealing - lower wages, possibly fewer job offers. On the other hand, the decision to keep daughters in school may be motivated by the high cost of preceding education to support the household. Indeed, schooling gives daughters a lot of time to participate in housework, whereas a job would make them less available. Furthermore, the conservative Costa Rican society may stigmatize young women who work, which may limit their opportunities. Further research is needed to disentangle these potential mechanisms.

Because the son is not rejected as a decision-maker, I can recover the bargaining function by mapping the estimated and structural parameters. The mapping is explained in detail in

Appendix 2.D.3. The son's bargaining function from equation 2.6 is[15]:

$$\rho_t = \kappa_1 + \underset{(0.223)}{1.146}\, w_t + \underset{(0.753)}{1.796}\, \ln w_p + \underset{(0.505)}{2.190}\, y, \quad \text{if } s^t = 0$$
$$\rho_t = \kappa_0 + \underset{(0.051)}{0.821}\left(\underset{(0.223)}{1.146}\, w_t + \underset{(0.753)}{1.796}\, \ln w_p + \underset{(0.505)}{2.190}\, y \right), \quad \text{if } s^t = 1$$

From the equations, the son's parents assign him 0.146 extra units from the household total income for an additional unit in his wage. He obtains 1.190 extra units for every unit of nonlabor income in the household. These additional units may appear considerable, but keep in mind that the teen's wage and the household's nonlabor income are modest compared with the parents' wage. In exchange, the parents transfer 0.1796 units (0.1 log point) for every 10% rise in their wages. Finally, if the son studies, he receives 82.1% of the total income he receives if he works.

Regarding the home production function, Table A13 shows the results. The daughter has a productivity parameter in the public good production of 0.045 units, while the sons have one of 0.038. This difference in productivity might imply an early-age work specialization where daughters are expected to work more in the household than their male counterparts. In contrast, sons work in the labor market instead. However, these results are not as significant as those from the descriptive statistics shown above with daughters providing more domestic hours per week than their male counterparts. The Cobb Douglas assumption on the production function could explain the brief discrepancy in productivity statistics. Because this technology has a substitution parameter equal to one. By relaxing this assumption, I might find a larger difference between daughters and sons. However, to do so, I need to find production shifters that work well with my sample.

1.8 Conclusion

Teens are key members of the household. However, the majority of analyses of them as decision-makers in households have only looked at consumption models. In this study, I show that they play a role in how households allocate time and money. First, I examine the effects of the Costa Rican conditional cash transfer *Avancemos* on the parents' and teens' time allocation. Because the transfer is endogenous to the household's decision-making, I estimate the marginal treatment effect of the teen's education decision, the father's labor supply, the mother's employment status, and the members' domestic work for those households that are unsure whether to apply or not. I find that when parents of daughters receive the transfer, they reduce their working hours and

[15] Standard errors are estimated using the Delta Method.

increase their leisure time, indicating a standard income effect. In households with a son, the only effects are on him, who is more likely to attend high school, and his mother, who decreases her labor market involvement. I explain that this gender difference arises from sons bargaining with their parents but not daughters. I provide a collective home model that allows for flexible time allocation between schooling, labor market work, and domestic work to investigate the existence of this bargaining. My findings reject the daughter, but not the son, as a decision-maker, which is consistent with the marginal treatment effect findings.

These findings have significant implications for public policy formulation. Because sons bargain with their parents, they consider the potential wage and the percentage of income allocated to them. Subsidies that are relatively low to encourage sons to stay in school may be ineffective since they do not compensate for the loss in resource sharing that he would receive if he worked. In addition, daughters complement their education with more domestic work, which lowers their opportunity costs to attend school. These two potential mechanisms could also lead to sons specializing in labor and daughters specializing in domestic work. This might explain why, in their adult ages, women have less household bargaining power and lower participation in the labor market. Public policies aimed at closing gender gaps must consider those gender roles starting at a young age in households.

There is plenty of room for future investigation. It would be interesting to learn who the teen steals bargaining power from. This, however, requires recognizing the mother and father as separate decision-makers. Even if the literature can easily be expanded to cover this concept, the estimation of such models has strong data requirements. It requires enough wage and labor supply for each decision-maker. Additionally, for two of the members an exogenous variation, such as a CCT, without any interaction with each other. The same research could be done on family members, such as grandparents, and other children living in the home. It would also be necessary to examine how much the work of teens, especially daughters, influences home production. Important gender differences resulting from work specialization can be highlighted by this analysis, which may also help to explain a significant portion of women's low labor participation rates.

Bibliography

Andresen, M. E. (2018). Exploring marginal treatment effects: Flexible estimation using stata. *The Stata Journal*, 18(1):118–158.

Ashraf, N., Bau, N., Low, C., and McGinn, K. (2020). Negotiating a better future: How interpersonal skills facilitate intergenerational investment. *The Quarterly Journal of Economics*, 135(2):1095–1151.

Attanasio, O. P. and Lechene, V. (2014). Efficient responses to targeted cash transfers. *Journal of political Economy*, 122(1):178–222.

Blundell, R., Chiappori, P.-A., Magnac, T., and Meghir, C. (2007). Collective labour supply: Heterogeneity and non-participation. *The Review of Economic Studies*, 74(2):417–445.

Blundell, R., Chiappori, P.-A., and Meghir, C. (2005). Collective labor supply with children. *Journal of political Economy*, 113(6):1277–1306.

Blundell, R. and Costa Dias, M. (2009). Alternative approaches to evaluation in empirical microeconomics. *Journal of Human Resources*, 44(3):565–640.

Blundell, R. and Powell, J. L. (2003). Endogeneity in nonparametric and semiparametric regression models. *Econometric society monographs*, 36:312–357.

Bonnal, L., Fougère, D., and Sérandon, A. (1997). Evaluating the impact of French employment policies on individual labour market histories. *The Review of Economic Studies*, 64(4):683–713.

Brown, C., Calvi, R., and Penglase, J. (2018). Sharing the pie: Undernutrition, intra-household allocation, and poverty. *Intra-Household Allocation, and Poverty (June 1, 2018)*.

Calvi, R. (2020). Why are older women missing in India? The age profile of bargaining power and poverty. *Journal of Political Economy*, 128(7):2453–2501.

Calvi, R., Penglase, J., Tommasi, D., and Wolf, A. (2021). The more the poorer? Resource sharing and scale economies in large families. *CEPR Discussion Paper No. DP15924*.

Carneiro, P., Heckman, J. J., and Vytlacil, E. J. (2011). Estimating marginal returns to education. *American Economic Review*, 101(6):2754–81.

Cherchye, L., De Rock, B., and Vermeulen, F. (2012). Married with children: A collective labor supply model with detailed time use and intrahousehold expenditure information. *American Economic Review*, 102(7):3377–3405.

Chiappori, P.-A. (1992). Collective labor supply and welfare. *Journal of political Economy*, 100(3):437–467.

Chiappori, P.-A. and Mazzocco, M. (2017). Static and intertemporal household decisions. *Journal of Economic Literature*, 55(3):985–1045.

Cornelissen, T., Dustmann, C., Raute, A., and Schönberg, U. (2016). From late to mte: Alternative methods for the evaluation of policy interventions. *Labour Economics*, 41:47–60.

Cunha, F., Heckman, J. J., Lochner, L., and Masterov, D. V. (2006). Interpreting the evidence on life cycle skill formation. *Handbook of the Economics of Education*, 1:697–812.

Dauphin, A., El Lahga, A.-R., Fortin, B., and Lacroix, G. (2011). Are children decision-makers within the household? *The Economic Journal*, 121(553):871–903.

De Hoyos, R., Rogers, H., and Székely, M. (2016). Out of school and out of work: Risk and opportunities for latin america's ninis. *Working Paper World Bank, Washington, DC.*

De Rock, B., Potoms, T., and Tommasi, D. (2022). Household responses to cash transfers. *Economic development and cultural change*, 70(2):625–652.

Del Boca, D., Flinn, C., and Wiswall, M. (2014). Household choices and child development. *Review of Economic Studies*, 81(1):137–185.

DiPasquale, D. and Glaeser, E. L. (1999). Incentives and social capital: Are homeowners better citizens? *Journal of urban Economics*, 45(2):354–384.

Dubois, P. and Rubio-Codina, M. (2012). Child care provision: Semiparametric evidence from a randomized experiment in mexico. *Annals of Economics and Statistics/ANNALES D'ÉCONOMIE ET DE STATISTIQUE*, pages 155–184.

Dunbar, G. R., Lewbel, A., and Pendakur, K. (2013). Children's resources in collective households: Identification, estimation, and an application to child poverty in Malawi. *American Economic Review*, 103(1):438–71.

García, S. and Saavedra, J. E. (2017). Educational impacts and cost-effectiveness of conditional cash transfer programs in developing countries: A meta-analysis. *Review of Educational Research*, 87(5):921–965.

Heckman, J. J. and Vytlacil, E. (2005). Structural equations, treatment effects, and econometric policy evaluation 1. *Econometrica*, 73(3):669–738.

Hernández Romero, K. (2016). ¿Cómo funciona Avancemos?: Mejores prácticas en la implementación de programas de transferencias monetarias condicionadas en América Latina y el Caribe. *BID*.

Instituto Mixto de Ayuda Social de Costa Rica (2022). Avancemos. Accessed: 22-02-2022.

Instituto Ncional de Estadísticas y Censos (2015). Canasta básica alimentaria - julio 2015. Accessed: 26-04-2022.

Instituto Ncional de Estadísticas y Censos (2022). Encuesta Nacional de Hogares. Accessed: 30-03-2020.

Jiménez-Fontana, P. (2015). Analysis of non-remunerated production in Costa Rica. *The Journal of the Economics of Ageing*, 5:45–53.

Keshavarz Haddad, G. (2017). Parents' decision on child labour and school attendance: Evidence from Iranian households. *Journal of Education and Work*, 30(6):612–631.

Lang Clachar, G., Soto Méndez, B., and Robalino, J. (2015). Efecto del programa "Avancemos" sobre el trabajo infantil en Costa Rica. Technical report.

Lundberg, S., Romich, J. L., and Tsang, K. P. (2009). Decision-making by children. *Review of Economics of the Household*, 7(1):1–30.

Mata, C. and Hernández, K. (2015). Evaluación de impacto de la implementación de transferencias monetarias condicionadas para educación secundaria en Costa Rica (Avancemos). *Revista de Ciencias Económicas*, 33(1):9–35.

Meza-Cordero, J. A. (2011). The effects of subsidizing secondary schooling: Evidence from a conditional cash transfer program in Costa Rica. *Unpublished Document, University of Southern California*.

Muñoz-Alvarado, J. A. (2016). Avancemos: Efectos sobre el abandono educativo en el hogar de personas beneficiarias. *Revista Electrónica Educare*, 20(1):53–74.

Reggio, I. (2011). The influence of the mother's power on her child's labor in Mexico. *Journal of development economics*, 96(1):95–105.

Sokullu, S. and Valente, C. (2022). Individual consumption in collective households: Identification using repeated observations with an application to Progresa. *Journal of Applied Econometrics*, 37(2):286–304.

Stern, N. (1986). On the specification of labour supply functions. In Blundell, R. W., Blundell, R., and Walker, I., editors, *Unemployment, search and labour supply*. Cambridge University Press.

Todd, P. E. and Wolpin, K. I. (2006). Assessing the impact of a school subsidy program in mexico: Using a social experiment to validate a dynamic behavioral model of child schooling and fertility. *American economic review*, 96(5):1384–1417.

Tommasi, D. (2019). Control of resources, bargaining power and the demand of food: Evidence from Progresa. *Journal of Economic Behavior & Organization*, 161:265–286.

Vermeulen, F. (2002). Collective household models: Principles and main results. *Journal of Economic Surveys*, 16(4):533–564.

1.A Reduced form results

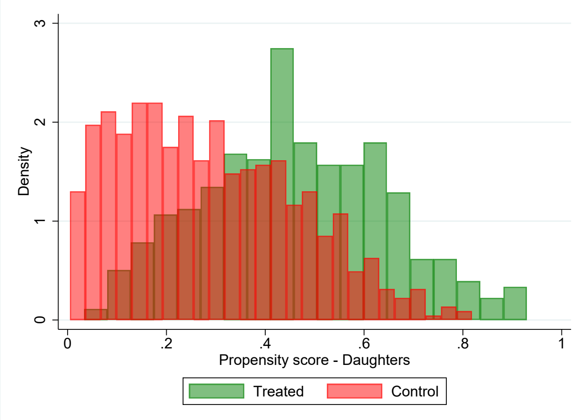
Table A1 shows the results of the first stage of the MTE estimation for the three samples: complete, daughters and sons. Figure A1 shows the graphs of the propensity scores. I estimate them using a probit.

Table A1: First stage: *Avancemos*' marginal effects

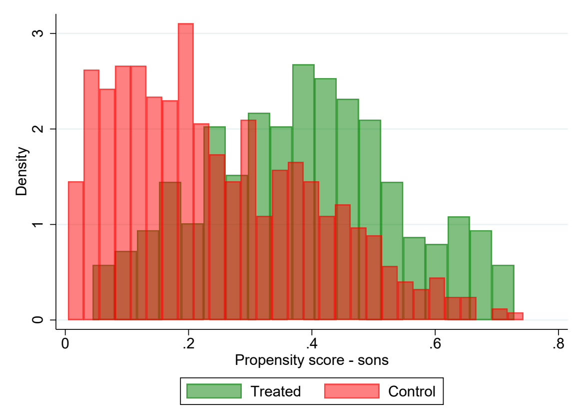
	Daughters	Sons
Percentage of households with *Avancemos*	2.138***	1.449***
	(0.271)	(0.256)
Father years of schooling	0.000128	0.00448
	(0.00678)	(0.00580)
Father high school or more diploma	-0.0646	-0.133**
	(0.0575)	(0.0500)
Father occupation mechanic	0.122***	0.0138
	(0.0328)	(0.0288)
Father occupation elemental	0.135***	0.0336
	(0.0338)	(0.0292)
Father occupation services and transport	-0.0588*	-0.0392
	(0.0297)	(0.0253)
Father insurance employed	-0.0506	-0.00675
	(0.0275)	(0.0237)
Mother years of schooling	-0.00363	-0.00646
	(0.00654)	(0.00622)
Mother high school or more diploma	-0.189**	-0.0890
	(0.0586)	(0.0514)
Teen age 18 or more	-0.239***	-0.221***
	(0.0363)	(0.0275)
Age father	0.000264	-0.00101
	(0.00229)	(0.00215)
Age mother	-0.000192	-0.000141
	(0.00293)	(0.00260)
One younger children	0.120***	0.178***
	(0.0363)	(0.0296)
Year effects	Yes	Yes
Geographical effects	Yes	Yes
Observations	1,090	1,357
R^2	0.146	0.127

Standard errors between parenthesis. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Baseline categories: no education diploma or school diploma for parents, different occupations (manager, research, technical and academic professors, staff and agriculture), different industries (mines and agriculture, finance, public administration, real state, teaching, social health, domestic and others), no insurance or non-employed insurance for the father, no younger children in the household.

Figure A1: Propensity scores for receiving *Avancemos*



(a) Daughters.



(b) Sons.

Table A2 shows the estimation results of observables for the MTE for the daughters' sample and Figure A2 shows the graphs of the estimated unobservable.

Table A2: *Avancemos'* effect on households for the daughters sample: MTE.

	Teen's schooling		Teen's domestic hours		Mother's employment		Mother's domestic hours		Father's market hours		Father's domestic hours	
	Baseline	Difference if treated	Baseline	Difference if treated	Baseline	Difference if treated	Baseline	Difference if treated	Baseline	Difference if treated	Baseline	Difference if treated
Father years of schooling	0.015*	-0.015*	-0.421*	0.327	0.001	0.003	0.009	1.518*	0.095	0.320	0.173	0.109
	(0.007)	(0.007)	(0.192)	(0.322)	(0.009)	(0.016)	(0.388)	(0.671)	(0.229)	(0.373)	(0.104)	(0.189)
Father high school or more diploma	-0.050	0.050	-0.190	-1.706	-0.027	-0.052	0.006	-11.123*	-3.338*	-3.037	-0.134	-0.881
	(0.055)	(0.055)	(1.459)	(2.236)	(0.071)	(0.124)	(2.962)	(5.496)	(1.691)	(3.132)	(0.887)	(1.476)
Father occupation elemental	-0.005	0.005	1.204	-2.119	0.161**	-0.182*	-4.863*	3.699	-1.760	0.272	0.368	-0.554
	(0.051)	(0.051)	(1.063)	(1.603)	(0.052)	(0.077)	(2.102)	(3.503)	(1.404)	(2.229)	(0.729)	(1.018)
Father occupation services and transport	-0.026	0.026	-0.155	1.206	0.005	0.078	-3.496*	1.839	1.449	2.129	-0.475	0.973
	(0.030)	(0.030)	(0.889)	(1.476)	(0.041)	(0.072)	(1.636)	(2.875)	(0.977)	(1.873)	(0.523)	(0.814)
Father insurance employed	0.002	-0.002	-0.489	1.132	-0.095*	0.074	2.151	-4.057	3.010**	3.172	0.029	0.319
	(0.035)	(0.035)	(0.893)	(1.313)	(0.039)	(0.062)	(1.580)	(2.763)	(0.982)	(1.639)	(0.457)	(0.728)
Mother years of schooling	0.025***	-0.025***	-0.160	0.399	0.021*	-0.012	-0.105	-0.386	-0.099	-0.480	0.332**	-0.385*
	(0.007)	(0.007)	(0.168)	(0.304)	(0.009)	(0.016)	(0.354)	(0.674)	(0.231)	(0.372)	(0.124)	(0.194)
Mother high school or more diploma	-0.097	0.097	0.063	0.437	0.112	0.002	-4.599	5.650	-2.231	3.936	0.980	-0.293
	(0.056)	(0.056)	(1.635)	(2.733)	(0.077)	(0.149)	(3.156)	(6.276)	(1.821)	(3.405)	(0.946)	(1.367)
Teen age 18 or more	-0.322***	0.322***	4.326***	-0.678	-0.118	0.065	-2.110	3.542	-3.222*	0.414	0.752	-1.443
	(0.054)	(0.054)	(1.290)	(2.342)	(0.060)	(0.104)	(2.408)	(4.815)	(1.386)	(2.973)	(0.681)	(1.174)
Age mother	-0.000	0.000	-0.003	0.046	-0.004	-0.004	0.237	0.001	-0.255*	0.096	0.041	-0.048
	(0.003)	(0.003)	(0.084)	(0.139)	(0.004)	(0.008)	(0.156)	(0.300)	(0.105)	(0.193)	(0.048)	(0.084)
One younger children	0.024	-0.024	0.425	-2.316	-0.072	0.039	9.052***	-2.625	0.065	-1.432	-0.033	-0.886
	(0.038)	(0.038)	(1.036)	(1.808)	(0.049)	(0.087)	(1.987)	(3.548)	(1.172)	(2.291)	(0.646)	(1.018)
Two or more younger children	-0.020	0.020	1.712	-3.905*	-0.125*	0.034	16.746***	-6.177	0.299	0.510	0.733	-1.457
	(0.050)	(0.050)	(1.128)	(1.954)	(0.054)	(0.096)	(2.369)	(4.105)	(1.458)	(2.558)	(0.751)	(1.144)
Constant	0.475**	0.525**	15.941***	-0.439	0.432	0.088	22.306**	0.760	65.499***	-10.036	-3.764	4.958
	(0.175)	(0.175)	(4.400)	(7.401)	(0.227)	(0.370)	(8.634)	(14.303)	(6.257)	(9.391)	(2.930)	(4.460)
Year and geographical effects	Yes		Yes		Yes		Yes		Yes		Yes	
P-value observable heterogeneity	0.0000		0.6466		0.0293		0.1857		0.4465		0.0481	
P-value no unobservable heterogeneity	0.991		0.9993		0.8920		0.9996		0.8150		0.8719	
Observations	1,090		1,090		1,090		1,090		1,090		1,090	

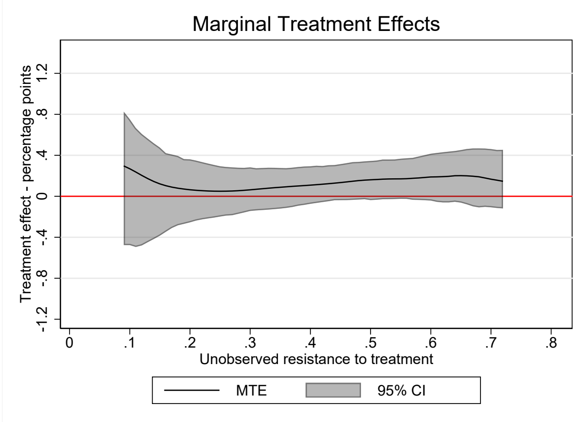
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Baseline categories: no education diploma for parents, different occupations (manager, research, technical and academic professors and staff), different industries (finance, public administration, real state, teaching, social health, domestic and others), no insurance, no younger children in the household and for the geographical variable it is living outside the Central Valley in a rural zone.

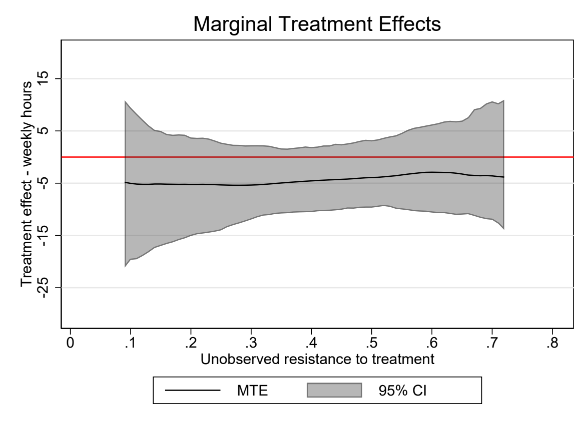
Omitted variables from the table because of their non significant effect: father's age, mother's high school diploma or more, indicator variable for father with a mechanic occupation.

The reported test statistic for observable heterogeneity tests for the joint significance of all elements in the column "Difference if treated", while the test for unobservable heterogeneity tests if MTEs differ with unobserved costs of treatment. I estimated the model in STATA with the user written command "mtefe" (Andresen, 2018).

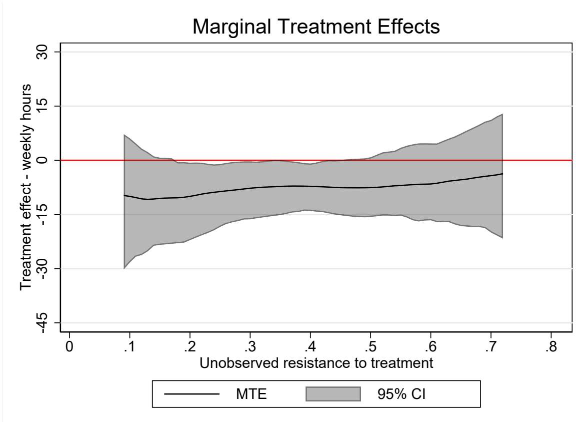
Figure A2: *Avancemos* on household with teen daughter: MTE.



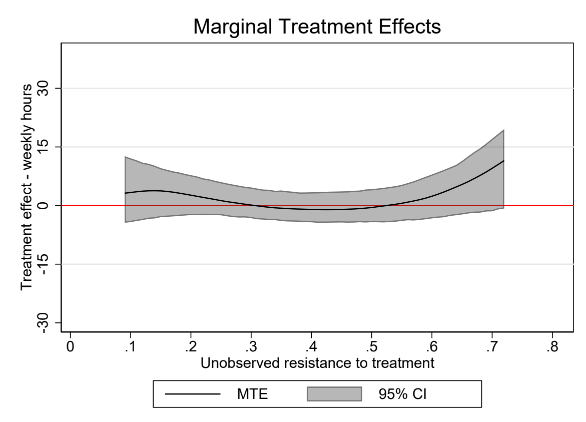
(a) Teen's schooling decision.



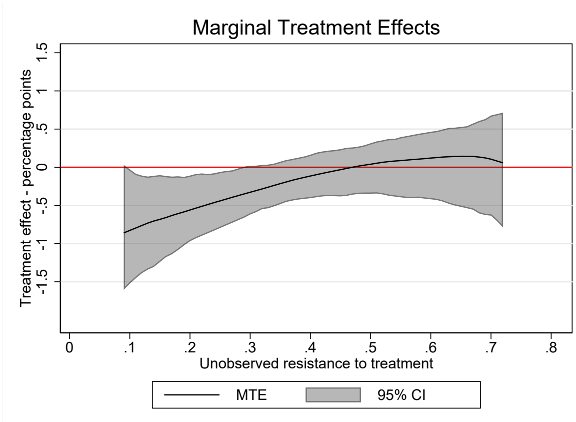
(b) Teen's domestic labor supply.



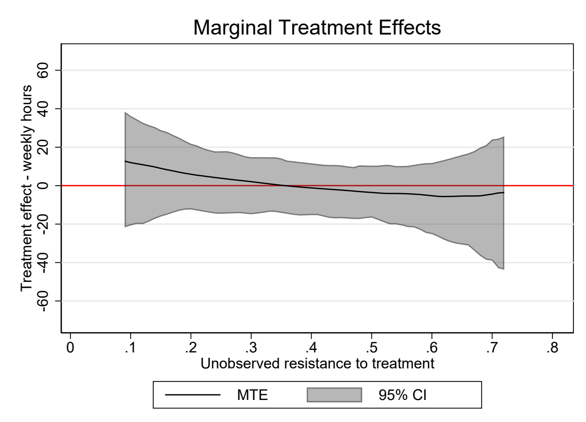
(c) Father's market labor supply.



(d) Father's domestic supply.



(e) Mother's employment status.



(f) Mother's domestic supply.

Table A3 shows the estimation results of observables for the MTE for the sons' sample and Figure A3 shows the graphs of the estimated unobservable.

Table A3: *Avancemos'* effect on households for the sons sample: MTE.

	Teen's schooling		Teen's domestic hours		Mother's employment		Mother's domestic hours		Father's market hours		Father's domestic hours	
	Baseline	Difference if treated	Baseline	Difference if treated	Baseline	Difference if treated	Baseline	Difference if treated	Baseline	Difference if treated	Baseline	Difference if treated
Father high school or more diploma	0.085	-0.085	-0.874	-1.523	0.019	-0.039	-0.165	-0.720	-3.323*	1.591	-0.480	0.567
	(0.055)	(0.055)	(0.760)	(1.325)	(0.070)	(0.155)	(2.712)	(6.043)	(1.642)	(3.302)	(0.795)	(1.672)
Father occupation elemental	-0.124**	0.124**	-0.228	0.832	-0.053	0.080	-0.028	-2.066	-2.973**	1.371	0.046	0.010
	(0.039)	(0.039)	(0.436)	(0.754)	(0.039)	(0.066)	(1.571)	(2.869)	(1.015)	(1.865)	(0.464)	(0.849)
Father occupation services and transport	-0.065*	0.065*	-0.162	1.050	0.051	-0.051	-2.255	0.206	3.125***	-0.004	-0.522	0.058
	(0.029)	(0.029)	(0.365)	(0.694)	(0.035)	(0.068)	(1.307)	(2.735)	(0.858)	(1.810)	(0.376)	(0.725)
Father insurance employed	0.010	-0.010	0.650	-1.014	0.046	-0.140*	0.277	1.078	2.520***	4.068**	0.364	-0.358
	(0.029)	(0.029)	(0.383)	(0.698)	(0.029)	(0.059)	(1.367)	(2.409)	(0.686)	(1.393)	(0.348)	(0.723)
Mother years of schooling	0.020**	-0.020**	0.051	0.089	0.025**	-0.016	-1.109**	1.154	-0.211	-0.044	0.289**	0.030
	(0.008)	(0.008)	(0.096)	(0.184)	(0.008)	(0.017)	(0.356)	(0.694)	(0.181)	(0.414)	(0.101)	(0.180)
Mother high school or more diploma	0.020	-0.020	-0.285	0.402	-0.066	-0.094	2.841	0.628	-0.519	2.873	-1.729*	1.002
	(0.056)	(0.056)	(0.790)	(1.754)	(0.067)	(0.134)	(2.768)	(5.752)	(1.579)	(3.479)	(0.868)	(1.729)
Teen age 18 or more	-0.414***	0.414***	-0.842	1.868	-0.077	-0.078	0.017	0.038	-1.325	5.110	-1.248	0.839
	(0.047)	(0.047)	(0.654)	(1.275)	(0.054)	(0.101)	(1.964)	(4.999)	(1.199)	(3.078)	(0.689)	(1.389)
Age father	0.003	-0.003	-0.032	-0.071	-0.004	0.003	0.138	-0.606**	-0.086	-0.292	-0.074*	0.030
	(0.003)	(0.003)	(0.036)	(0.061)	(0.003)	(0.006)	(0.117)	(0.221)	(0.067)	(0.154)	(0.034)	(0.064)
Age mother	-0.005	0.005	0.036	-0.062	-0.007	0.004	0.333*	0.048	-0.104	0.130	-0.004	-0.010
	(0.003)	(0.003)	(0.043)	(0.076)	(0.004)	(0.007)	(0.132)	(0.272)	(0.086)	(0.192)	(0.038)	(0.079)
One younger children	0.044	-0.044	1.381*	-1.246	0.057	-0.098	4.531*	4.229	0.031	-2.105	0.527	0.419
	(0.043)	(0.043)	(0.604)	(1.218)	(0.057)	(0.105)	(2.151)	(4.749)	(1.268)	(2.708)	(0.590)	(1.297)
Two or more younger children	-0.052	0.052	1.166*	-1.334	-0.099	-0.015	12.808***	2.314	0.357	-2.991	1.125	-0.258
	(0.042)	(0.042)	(0.515)	(1.050)	(0.053)	(0.102)	(2.139)	(4.595)	(1.182)	(2.748)	(0.619)	(1.219)
Constant	0.697***	0.303	2.879	4.693	0.701***	-0.725	18.698*	24.711	59.618***	2.548	4.338*	-2.609
	(0.165)	(0.165)	(1.992)	(4.365)	(0.183)	(0.448)	(7.419)	(17.844)	(4.218)			
Year and geographical effects	Yes		Yes		Yes		Yes		Yes		Yes	
P-value observable heterogeneity	0.0000		0.0662		0.0142		0.0034		0.3028		0.9556	
P-value no unobservable heterogeneity	0.9980		0.9957		0.8967		0.9988		1.000		0.9333	
Observations	1,357		1,357		1,357		1,357		1,357		1,357	

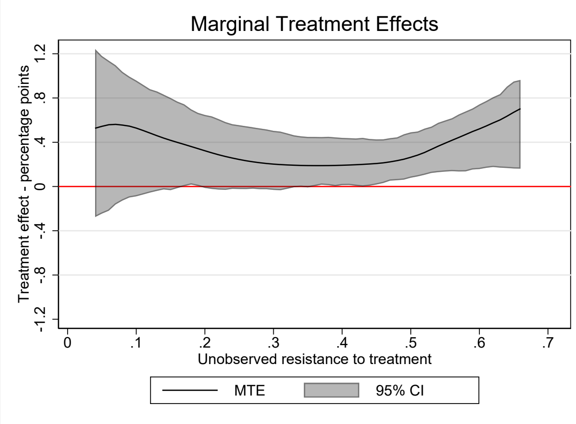
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Baseline categories: no education diploma for parents, different occupations (manager, research, technical and academic professors and staff), different industries (finance, public administration, real state, teaching, social health, domestic and others), no insurance, no younger children in the household and for the geographical variable it is living outside the Central Valley in a rural zone.

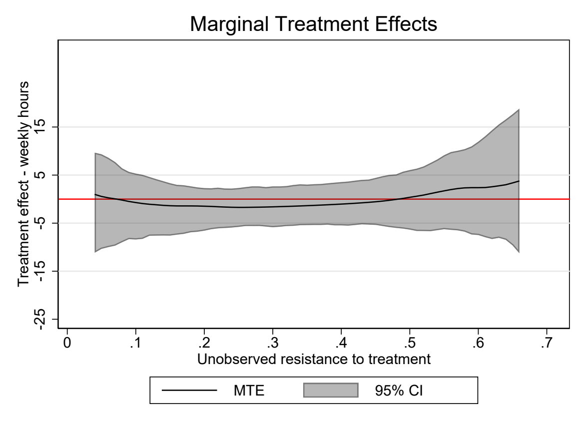
Omitted variables from the table because of their non significant effect: father's years of schooling and indicator variable for father with a mechanic occupation.

The reported test statistic for observable heterogeneity tests for the joint significance of all elements in the column "Difference if treated", while the test for unobservable heterogeneity tests if MTEs differ with unobserved costs of treatment. I estimated the model in STATA with the user written command "mtefe" (Andresen, 2018).

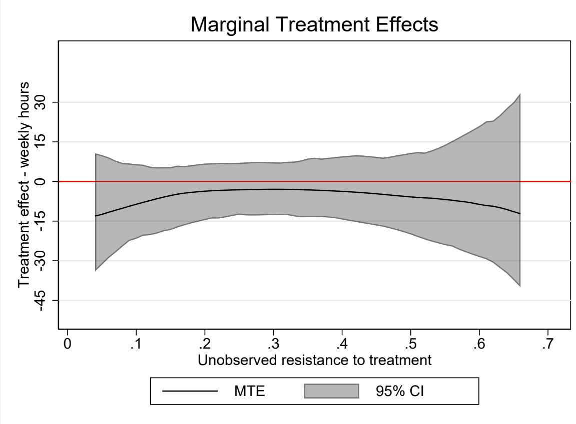
Figure A3: *Avancemos* on household with teen son: MTE.



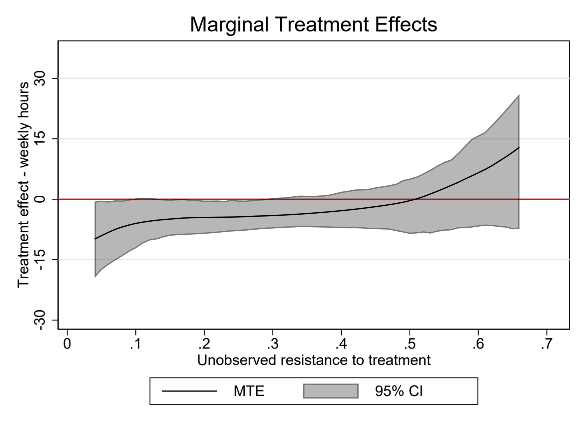
(a) Teen's schooling decision.



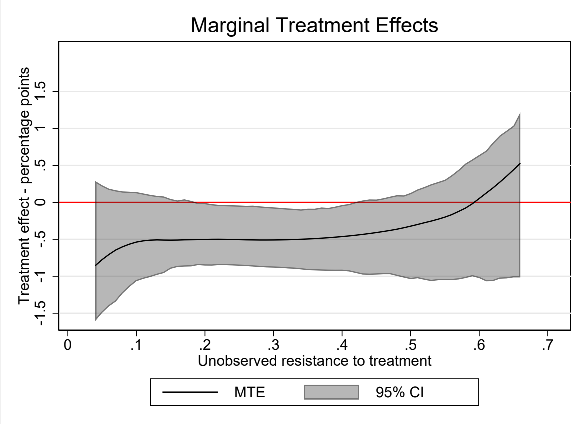
(b) Teen's domestic labor supply.



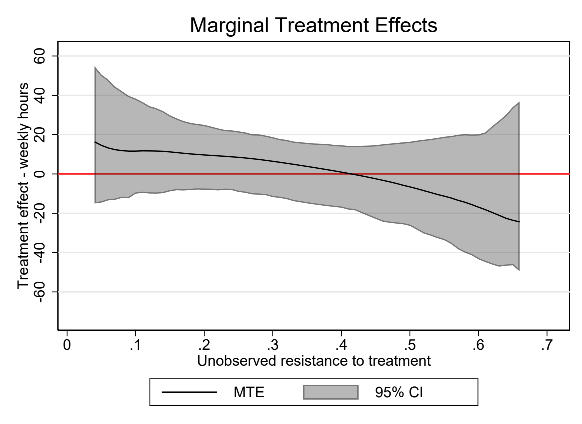
(c) Father's market labor supply.



(d) Father's domestic supply.



(e) Mother's employment status.



(f) Mother's domestic supply.

1.B Extensive version of models and their restrictions

1.B.1 Unitary model

In this section, I present the restrictions from the unitary model A17. The model derives two restrictions immediately. First, the "income pooling" property:

$$\frac{dm^p}{dw^t} = \frac{dm^p}{dy}, \quad \text{when teen works} \tag{UR1}$$

which means that a change in w^t can only have an income effect on the parents' market supply. On the other hand, if the teen goes to school, this effect is zero:

$$\frac{dm^p}{dw^t} = 0, \quad \text{when teen goes to school} \tag{UR2}$$

Regarding the teen's schooling decision, it depends on the difference between the household's (indirect) utility when he goes to school (denote V^s) or when she works (V^w):

$$s^t = 1 \iff V^s(w^p, y) \geq V^w(w^p, w^t + y)$$

The frontier is characterized by

$$V^w(w^p, y + \gamma(w^t + y)) = V^s(w^p, y)$$

which implies the following condition

$$\frac{d\gamma}{dw^p} = m^p_{school} - m^p_{work} + m^p_{school}\frac{d\gamma}{dy} \tag{UR3}$$

The last term on the right-hand side corresponds to a standard income effect. The difference $m^p_{school} - m^p_{work}$ corresponds to the effect on the teen's schooling decision cost due to a reduction of parents' labor supply. Restrictions UR1, UR2 and UR3 translate to the empirical model to equations U1 and U2 presented in the paper:

$$\begin{aligned} A_t &= A_y \\ a_t &= 0 \end{aligned} \tag{U1}$$

$$\begin{aligned} (1+\gamma_y)(a_y - A_y) &= 0 \\ A_y\gamma_p &= (1+\gamma_y)a^*_p - A^*_p \end{aligned} \tag{U2}$$

1.B.2 Collective model

The collective model 2.2 provides theoretical restrictions that are imposed in the estimation process. As presented by Blundell et al. (2007), the teen's extensive margin decision to go to school or work produces two different outcomes for the household, specifically her parents.
If the teen works, her utility is $U^t(0, C, f^K(h^p, h^t))$, which defines a fix level of utility:

$$U^t(0, C, \bar{f}^K) = \bar{u}^t(w^p, w^t, y)$$

Solving for consumption C^t, the optimal consumption is:

$$C^t = V^t\left[\bar{u}^t(w^p, w^t, y)\right] = \rho^t(w^p, w^t, y|\bar{f}^K)$$

where V^t is the inverse mapping of $U^t(0, \cdot)$. Pareto efficiency is equivalent to the parents' labor supply decision being the solution to the following program:

$$\begin{aligned} \max_{m^p, C^p} \; & U^p(1 - m^p, C^p, \bar{f}^K) \\ C^p &= w^p m^p + w^t + y - \rho^t(w^p, w^t, y|\bar{f}^K) \\ 0 < & m^p \leq 1 \end{aligned} \tag{A1}$$

This generates a labor supply of the form:

$$m^p(w^p, w^t, y) = M^p[w^p, w^t + y - \rho^t(w^p, w^t, y|\bar{f}^K)] \tag{A2}$$

where M^p is the Marshallian labor supply.

In the case the teen does not participate in the labor market and instead goes to school, her utility is $U^t(1, C, \bar{f}^K)$ and:

$$U^t(1, C, \bar{f}^K) = \bar{u}^t(w^p, w^t, y) = (V^t)^{-1}[\rho^t(w^p, w^t, y|\bar{f}^K)]$$

The teen's consumption can be obtained by inverting the previous equation

$$C^t = W^t[(V^t)^{-1}(\rho^t(w^p, w^t, y|\bar{f}^K))] = F(\rho^t(w^p, w^t, y|\bar{f}^K))$$

where W^t is the inverse of the mapping $U^t(1, \cdot)$ and $F = W^t \circ (V^t)^{-1}$ is increasing.

The parents' program follows as before and leads to the following labor supply:

$$m^p(w^p, w^t, y) = M^p[w^p, y^* - F(\rho^t(w^p, w^t, y|\bar{u}^K))] \tag{A3}$$

where M^p is the Marshallian labor supply. Lastly, to model the teen's schooling decision, the participation frontier, L, is a set of wages and non-labor income bundles $(w^p, w^t, y) \in L$, for which the teen is indifferent to attend high school or not. This implies that there exists a reservation wage for the teen w_R^t such that:

$$\forall (w^p, w^t, y) \in L, \quad \rho^t(w^p, w^t, y|\bar{u}^K) - F(\rho^t(w^p, w^t, y|\bar{u}^K)) = w_R^t \tag{A4}$$

Using a shadow wage condition to parametrize L, as define in equation (A4). The teen works if and only if:

$$w_R^t > \gamma(w^p, y)$$

for some function γ that describes the frontier.

The extended version of the empirical specification of the collective model presented in the paper and the derivation of its restrictions goes as follows. First, I focus on the first-stage allocation of the household's non-labor income y through the conditional sharing functions $\rho^i(w^p, w^t, y|\bar{f}^K)$. The first-stage household maximization boils down to:

$$\max_{\rho^p, \rho^t, f^K} \; \lambda^p(w^p, w^t, y)v^p(w^p, \rho^p, f^K) + \lambda^t(w^p, w^t, y)v^t(w^t, \rho^t, f^K) \tag{A5a}$$

$$s.t. \begin{cases} \lambda^p(w^p, w^t, y) + \lambda^t(w^p, w^t, y) = 1 & \text{(A5b)} \\ \rho^p + \rho^t + g^K(w^p, w^t)f^K = y^* & \text{(A5c)} \end{cases}$$

Assuming an interior solution with μ as the Lagrange multiplier, the first order conditions of the associated Lagrangian are:

$$\frac{\partial \mathcal{L}}{\partial \rho^p} = \lambda \frac{\exp(\theta_\rho^p w^p)}{\theta_\rho^p} \theta_\rho^p - \mu = 0$$

$$\frac{\partial \mathcal{L}}{\partial \rho^t} = (1 - \lambda)\theta_\rho^t - \mu = 0$$

$$\frac{\partial \mathcal{L}}{\partial f^K} = \lambda \frac{\exp(\theta_\rho^p w^p)}{\theta_\rho^p} \frac{\theta_K^p}{f^K} + (1 - \lambda)\frac{\theta_K^t}{f^K} - \mu g^K(w^p, w^t) = 0$$

$$\frac{\partial \mathcal{L}}{\partial \mu} = y^* - \rho^p - \rho^t - g^K(w^p, w^t)f^K = 0$$

From these equations, one obtains the Bowen-Lindahl-Samuelson condition for the optimal provision of public goods inside the household. Then using Roy's identity it is possible to recover the parents' Marshallian labor supply as presented in the paper. The restrictions for the collective model can be separated into two sets. The first set of restrictions comes from the labor participation of the teen. Blundell et al. (2007) show that equation (A4) has a unique solution for w_R^t under the following sufficient condition

$$\forall (w^p, w^t, y), |[1 - F'(\rho(w^p, \gamma(w^p, y), y))]g^K(w^p, w^t) = \frac{1}{\overline{f}^K}\left(\frac{\theta_K^p}{\theta_\rho^p} + \frac{\theta_K^t}{\theta_\rho^t}\right)(w^p, w^t, y)| < 1$$

which in my setting implies the following restriction:

$$|(1 - \theta_\rho^t)\psi_t| < 1 \tag{A6}$$

In this case, whenever $m^p > 0$, γ is characterized by the following equation:

$$\forall (w^p, y) \in L, \;\; \rho(w^p, \hat{m}^t(w^p, y), y) - F(\rho(w^p, \gamma(w^p, y), y)) = \gamma(w^p, y) \tag{A7}$$

which implies the following two conditions:

$$\psi_y + \gamma_y \psi_t = \frac{\gamma_y}{1 - \theta_\rho^t} \tag{CR1}$$

$$\psi_p = \frac{\gamma_{w^p}}{\gamma_y}\psi_y \tag{CR2}$$

The next set of restrictions is in the parents' labor supply. If the teen works, for any $(w^p, w^t, y) \in P$ such that $m^p(w^p, w^t, y) > 0$:

$$\frac{1 - \psi_t}{1 - \psi_y} = \frac{m_{w^t}^p}{m_y^p} = A(w^p, w^t, y)$$

In the case the teen goes to school:

$$\frac{-F'\psi_t}{1 - F'\psi_y} = \frac{m_{w^t}^p}{m_y^p} = B(w^p, w^t, y)$$

Both equations can be rearranged as:

$$-\psi_t + A\psi_y = A - 1 \tag{CR3}$$

$$-\psi_t + B\psi_y = \frac{B}{F'} \tag{CR4}$$

Aggregating restrictions (CR1), (CR2), (CR3) and (CR4) gives the two restrictions C1 presented in the paper:

$$\frac{A_t - a_t}{A_y - a_y} = -\frac{1}{\gamma_y}, \quad \frac{A_p^* - a_p^*}{A_y - a_y} = \frac{\gamma_p}{\gamma_y} \tag{C1}$$

1.B.3 Recovering structural parameters

If the data do not reject the teen as a decision maker, I can recover the sharing function of the household as done by Blundell et al. (2007). On any point of the frontier, the four restrictions above, (CR1), (CR2), (CR3) and (CR4), create a non-linear system of equations in the unknowns $(\psi_p, \psi_t, \psi_y, F')$. With some algebra, one obtains the following equation in F':

$$(\gamma_y ba - 1 + a - \gamma_y b)(F')^2 + (-b + 1 - 2\gamma_y ba + \gamma_y a - a)F' + b + \gamma_y ba = 0$$

where $a = a(w^p, y) = A[w^p, \gamma(w^p, y), y]$ and likewise for b. Blundell et al. (2007) show that if there is a solution to this quadratic equation that satisfies equation (A6), then the sharing rule is identified. This solution is such that:

$$F'(\rho^t(w^p, w^t, y|\bar{u}^K)) = \theta_\rho^t(w^p, y)$$

and (ψ_p, ψ_t, ψ_y) are recovered with the following equations (rewritten from the restrictions above):

$$\psi_t[w^p, \gamma(w^p, y), y] = K(w^p, y) = \frac{b}{(a-b)}\left(a - 1 - \frac{a}{\theta_\rho^t(w^p, y)}\right)$$
$$\psi_p[w^p, \gamma(w^p, y), y] = L(w^p, y) = \frac{\gamma_p}{(a-b)\gamma_y}\left(a - 1 - \frac{b}{\theta_\rho^t(w^p, y)}\right)$$
$$\psi_y[w^p, \gamma(w^p, y), y] = M(w^p, y) = \frac{1}{(a-b)}\left(a - 1 - \frac{b}{\theta_\rho^t(w^p, y)}\right)$$

Then, from the mapping between the structural Marshallian labor supply and its reduced form equation, I recover the last parameters:

$$\theta_\rho^p = \frac{A_y}{1 - \psi_y}$$
$$\theta_w^p = A_p + \theta_\rho^p \psi_p$$

Lastly, it is important to remind that functions ρ and F are identified up to a constant on the teen schooling participation.

1.C Wages' imputation

I impute teens' wages using a different sample of the Costa Rican National Household Survey from 2011 to 2019. The sample consists of individuals aged 15 to 25 years old with a school diploma and without a high school diploma who is the child of the head of the household. I impute men's and women's wages separately. The men's sample consists of 15,751 individuals where 4,824 are employed. For the women's sample, there are 1,397 employed women out of 11,810 observations. I impute using a Heckman two-step selection procedure with the following Mince equation for wages:

$$w^i = \alpha_0 + \alpha_1 \text{age}^i + \alpha_2 s^i + \mathbf{X}'\mathbf{A} + u_w^i \tag{A8}$$

where i is an individual, s^i is years of schooling and $\mathbf{X}$ is a vector of geographical and year effects. For the participation equation, I include as extra covariates demographic variables of the individual and household characteristics such as the number of children in the household, the head of the household's age and years of schooling. Table A1 shows the results of the estimation. For the women, Table A2 shows the results. Figure A5 shows the comparison between the observed and predicted values for both imputations.

Table A4: Men's wage imputation results

	Dependent variable:	
	Employed	Log Hourly wage rate
Age	0.268***	0.058***
	(0.005)	(0.011)
Years of schooling		0.035***
		(0.011)
One child in household	−0.033	
	(0.037)	
Two children	−0.033*	
	(0.037)	
Three children	0.068	
	(0.046)	
Four or more children	−0.033	
	(0.053)	
Years of schooling head of household	−0.039***	
	(0.004)	
Age head of household	−0.007***	
	(0.001)	
Constant	−5.029***	−0.425
	(0.112)	(0.292)
Year and geographical effects	Yes	Yes
Observations	15,751	4,824
R^2		0.076
Adjusted R^2		0.073
Log Likelihood	-6,833.110	
Akaike Inf. Crit.	13,704.220	
ρ		0.377
Inverse Mills Ratio		0.176*** (0.067)

*p<0.1, **p<0.05,***p<0.01.

Baseline categories: different occupations (manager, research, technical and academic professors and staff), different industries (finance, public administration, real state, teaching, social health, domestic and others), spouse or another relationship in the household, working in a firm with less than 10 employees and for the geographical variable it is living outside the Central Valley in a rural zone.

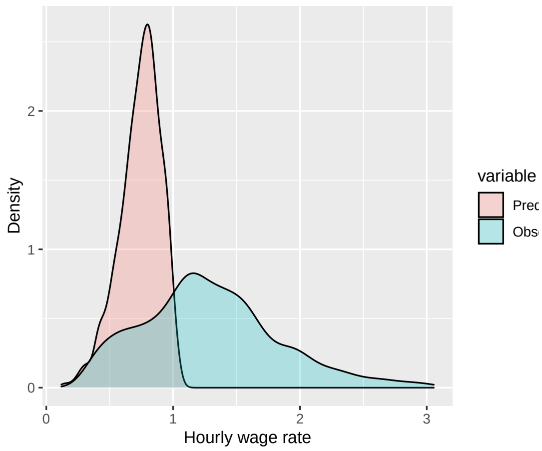
Table A5: Women's wage imputation results

	Dependent variable:	
	Employed	Log Hourly wage rate
Age	0.258***	0.074*
	(0.007)	(0.043)
Years of schooling		0.025**
		(0.010)
One child in household	0.064	
	(0.056)	
Two children	0.021	
	(0.058)	
Three children	0.121*	
	(0.068)	
Four or more children	0.207**	
	(0.079)	
Years of schooling head of household	−0.012**	
	(0.006)	
Age head of household	−0.004*	
	(0.002)	
Constant	−5.962***	−1.028
	(0.173)	(1.199)
Year and geographical effects	Yes	Yes
Observations	11,810	1,397
R^2		0.050
Adjusted R^2		0.039
Log Likelihood	-2,872.722	
Akaike Inf. Crit.	5,783.444	
ρ		0.532
Inverse Mills Ratio		0.312 (0.224)

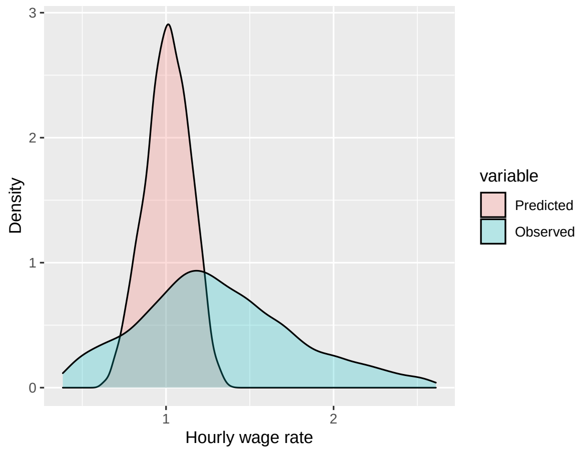
*p<0.1, **p<0.05, ***p<0.01.

Baseline categories: different occupations (manager, research, technical and academic professors and staff), different industries (finance, public administration, real state, teaching, social health, domestic and others), spouse or another relationship in the household, working in a firm with less than 10 employees and for the geographical variable it is living outside the Central Valley in a rural zone.

Figure A4: Imputation wages



(a) Women.



(b) Men.

1.D Control function approach

For non-labor income, I define as the difference between the household's total income and its labor income. The sample I use is the same as presented in the paper. To obtain the residuals I use in the control function approach I estimate the following regression:

$$y = \alpha_0^y + \alpha_1 \mathrm{IV}_i + \alpha_2 \mathrm{IV}_i^2 + \mathbf{X}'\mathbf{A} + u_{yi} \tag{A9}$$

where i is a household, IV is an instrument and X is a vector of covariates containing demographics and geographical and year effects. The instrument I use is the amount the household receives if it benefits from *Avancemos*. Table A3 show the results of the estimation.

Table A6: Non-labor income imputation results

	Dependent variable:
	non-labor income
IV	-1.085^{*}
	(0.591)
IV square	0.094
	(0.061)
Age teen	1.617^{***}
	(0.432)
Female teen	-2.274^{**}
	(1.124)
Age father	0.075
	(0.102)
Age mother	-0.281^{**}
	(0.127)
Father's years of schooling	0.307
	(0.285)
Mother's years of schooling	0.985^{***}
	(0.293)
Father high school diploma or more	4.902^{**}
	(2.308)
Mother high school diploma or more	7.269^{***}
	(2.365)
One kid	−0.207
	(1.469)
Two or more kids	3.520^{**}
	(1.685)
Constant	24.132^{***}
	(8.599)
Year and geographical effects	Yes
Observations	2,447
Log Likelihood	-11,550.520
Akaike Inf. Crit.	23,149.040

*p<0.1, **p<0.05,***p<0.01.
Baseline categories: positive rent income, household of 3 members (father, mother and teen) and for the geographical variable it is living outside the Central Valley in a rural zone.

1.E Structural results

In this section I show the results of the structural model. As explained in the paper, I estimate an unrestricted model, the unitary model and the collective model. I show three the results for each estimation for the sons and daughters samples.

Table A7: Estimates unrestricted model - daughters

	Parents weekly labor hours				**Teen school decision**	
	Teen no school		Teen school			
	Coef	**SE**	**Coef**	**SE**	**Coef**	**SE**
Hourly wage teen	9.478	(8.179)	11.979	(6.610)	-1.336	(0.262)
Hourly wage parents	-16.763	(5.995)	-15.406	(2.512)	0.479	(0.131)
Non-labor income	26.685	(19.408)	38.098	(7.175)	0.892	(0.390)
Intercept	34.042	(20.134)	22.761	(7.482)	0.321	(0.419)
Control function	-18.759	(18.842)	-22.988	(7.260)	-0.860	(0.390)
N	148		942		1,090	

The variable "Control function" refers to the residuals of the regression on non-labor income.

Missing values for some variables are due to the restrictions impose in the estimation from the model.

Table A8: Estimates unitary model - daughters

	Parents weekly labor hours				**Teen school decision**	
	Teen no school		Teen school			
	Coef	**SE**	**Coef**	**SE**	**Coef**	**SE**
Hourly wage teen	29.097	(5.517)			-1.451	(0.245)
Hourly wage parents	-26.282	(4.400)			0.439	(0.130)
Non-labor income					0.767	(0.387)
Intercept	23.238	(9.046)	36.268	(5.795)	0.539	(0.397)
Control function	-20.330	(7.573)	-15.155	(5.746)	-0.724	(0.386)
N	148		942		1,090	

The variable "Control function" refers to the residuals of the regression on non-labor income.

Missing values for some variables are due to the restrictions impose in the estimation from the model.

Table A9: Estimates collective model - daughters

	Parents weekly labor hours				**Teen school decision**	
	Teen no school		Teen school			
	Coef	**SE**	**Coef**	**SE**	**Coef**	**SE**
Hourly wage teen			7.211	(7.783)	-1.183	(0.420)
Hourly wage parents			-14.606	(2.566)	0.446	(0.156)
Non-labor income	24.450	(23.936)	36.279	(7.350)	1.269	(0.647)
Intercept	32.456	(21.782)	26.682	(7.069)	-0.155	(0.813)
Control function	-14.984	(23.217)	-21.461	(7.562)	-1.234	(0.637)
N	148		942		1,090	

The variable "Control function" refers to the residuals of the regression on non-labor income.
Missing values for some variables are due to the restrictions impose in the estimation from the model.

Table A10: Estimates unrestricted model - sons

	Parents weekly labor hours				**Teen school decision**	
	Teen no school		Teen school			
	Coef	**SE**	**Coef**	**SE**	**Coef**	**SE**
Hourly wage teen	3.461	(4.468)	24.535	(9.524)	-1.804	(0.220)
Hourly wage parents	-5.946	(3.841)	-12.117	(2.388)	0.745	(0.109)
Non-labor income	21.169	(12.673)	21.504	(6.160)	1.019	(0.290)
Intercept	33.432	(14.387)	23.054	(9.595)	0.503	(0.354)
Control function	-12.737	(12.605)	-12.644	(6.335)	-1.195	(0.295)
N	342		1,015		1,357	

The variable "Control function" refers to the residuals of the regression on non-labor income.
Missing values for some variables are due to the restrictions impose in the estimation from the model.

Table A11: Estimates unitary model - sons

	Parents weekly labor hours				**Teen school decision**	
	Teen no school		Teen school			
	Coef	**SE**	**Coef**	**SE**	**Coef**	**SE**
Hourly wage teen	17.480	(3.556)			-1.910	(0.214)
Hourly wage parents	-15.322	(3.037)			0.707	(0.109)
Non-labor income					0.933	(0.294)
Intercept	26.650	(7.376)	45.869	(4.149)	0.716	(0.351)
Control function	-9.498	(4.524)	-10.096	(3.963)	-1.102	(0.300)
N	342		1,015		1,357	

The variable "Control function" refers to the residuals of the regression on non-labor income.

Missing values for some variables are due to the restrictions impose in the estimation from the model.

Table A12: Estimates collective model - sons

	Parents weekly labor hours				**Teen school decision**	
	Teen no school		Teen school			
	Coef	**SE**	**Coef**	**SE**	**Coef**	**SE**
Hourly wage teen			21.117	(7.528)	-1.846	(0.218)
Hourly wage parents			-12.018	(2.238)	0.747	(0.108)
Non-labor income	26.709	(6.478)	17.911	(5.555)	0.911	(0.288)
Intercept	27.014	(7.752)	29.970	(7.902)	0.665	(0.351)
Control function	-18.250	(6.873)	-9.011	(5.797)	-1.086	(0.294)
N	342		1,015		1,357	

The variable "Control function" refers to the residuals of the regression on non-labor income.

Missing values for some variables are due to the restrictions impose in the estimation from the model.

Table A13: Estimates home production

Household production function				
	Coef	**SE**	**Coef**	**SE**
Production parameter				
Teen	0.045	(0.002)	0.038	(0.002)
Parents	0.955	(0.002)	0.962	(0.002)
Sample	Daughters		Sons	
N	1,090		1,357	

Standard errors computed with the Delta Method.

For the daughters' sample, the estimates are obtain from an OLS regression. For the sons, the estimates are from the maximization of the likelihood presented in the paper.

Chapter 2

You are the father! Effects of Costa Rica's Responsible Paternity Law on families

Abstract

Costa Rica's Responsible Paternity Law made it easier for unmarried women to declare the father of their newborn child and thus obtain child monetary support. This paper assesses the impact of the law on household decisions. I estimate the law's effects using the law as a natural experiment and a fuzzy differences-in-differences setting. I find that the law had a negative impact on male labor participation as well as female and male weekly labor supply. Using a collective household model with matching, I argue that the law strengthens women's bargaining power in household decision-making. This has two consequences: a couple selection effect and an intra-household allocation effect. Structural estimates show that both effects exist in households. These findings demonstrate how child-related laws help us better understand household formation and decision-making.

2.1 Introduction

Gender inequality affects women's daily lives in a variety of ways, including poverty, the labor market, and the marriage market. According to Wodon and De La Briere (2018), gender equality in earnings would increase human capital wealth by 21.7 percent and total wealth by 14 percent globally. Several studies have found that maternity is a contributing factor to gender inequality in households and the labor market. The Organisation for Economic Co-operation and Development (2012) mentions how large gender differences exist because women continue to bear the burden of unpaid domestic tasks such as childcare and housework.

Important gender inequality exists in the household, where the decision process between husband and wife affects the latter's well-being. Some policies help improve their situation, for example, Chiappori et al. (2017) study how the changes in alimony laws in Canada affect household labor decisions. They show how different female labor supplies were for those households before and after the law was approved. Goussé and Leturcq (2022) show how different levels of protection upon separation affect cohabited couples' labor supply. Yet, there is little evidence on how paternity laws affecting children have side effects on their mothers, the main caretaker.

In this paper, I study the effects of the Responsible Paternity Law (paternity law hereafter) of 2001 in Costa Rica on household formation and labor outputs. This law was enacted to ensure that all children have a registered father by allowing non-married mothers to automatically register the father of their newborn child. The main effect of the paternity law on women is the ability to seek monetary child support. I find two sets of results. First, by employing fuzzy differences-in-differences, the law reduced the labor supply for men and women who cohabit together. Second, I find this effect stems from a greater outside option for women in case of a potential pregnancy. This is reflected in a couple selection effect, in which women are more likely to remain single or cohabit rather than marry, as well as an intra-household bargaining effect, in which women enjoy a larger share of household resources.

To obtain empirical evidence of the paternity law, I use a sample of single, cohabited, and married individuals from the Costa Rican *Encuesta de Hogares de Propósitos Múltiples* from 1997 to 2009. For the first set of results, I use the fuzzy differences-in-differences framework as presented by De Chaisemartin and d'Haultfoeuille (2018). I define treated individuals as those who had a child after 2002, once the paternity law was in place. I find an 8% decrease in male labor participation and an average decrease of 5.5 weekly labor hours for women and 4.5 for men.

I use a collective model with matching, as presented by Choo and Seitz (2013), to explain couple formation and intra-household effects of the paternity law. This model proposes a simultaneous decision for men and women to form or not a household and the intra-household allocation of resources and consumption. Estimates show a positive effect of the law on the probability of a woman being single or cohabiting, while it decreases the probability of her getting married. The inverse happens with the man. If a man and woman decide to cohabit, the paternity law has an intra-household effect. It increases woman's bargaining power in household decision-making, allowing them to decrease their labor supply and enjoy more leisure. I do not find an intra-household effect on married couples.

This paper relates to the empirical literature on collective household models. Introduced by Chiappori (1992), collective models assume households behave according to cooperative bargaining between its decision-makers, primarily the father and mother. A review of the literature is presented in Chiappori and Mazzocco (2017). Recently, some papers started including a matching framework, allowing a link of the household bargaining function with couple formation decisions. For example, Choo and Seitz (2013) argue that the household bargaining function is determined in a previous stage decision where both potential spouses consider the marital gains to relative choices. I use their setting to estimate the effect of the paternity law on Costa Rican households. The novelty relies upon obtaining evidence that children's related laws also affect the parents' decisions.

My paper relates to the literature on laws affecting household behavior. Most papers focus on divorce laws. Reynoso (2018) studies the effects of introducing unilateral divorce in the United States. She finds that unilateral divorce increases assortative matching among newlyweds. Then, by using a life cycle model of marriage, labor supply, consumption, and divorce she finds that new-form couples share more socioeconomic backgrounds and that women are more likely to remain single. Goussé and Leturcq (2022) show how different levels of protection upon separation affect cohabited couples' labor supply in Canada. They find that eligibility for a regime making cohabiting partners equal to married partners increases men's labor supply and earnings and decreases women's while eligibility for a regime allowing for post-separation transfers between ex-partners decreases women's earnings only. My paper contributes to this literature by providing additional evidence on how laws providing similar rights to non-married couples as those who are married improve women's leisure consumption and welfare in case of separation.

Lastly, related to the literature on paternity laws, Rossin-Slater (2017) studies the effect of a paternity law on the US. She obtains that there is a decrease in the marriage rate due to lowering the cost of legal paternity establishment. Ekberg et al. (2013) use Swedish data to study the effect of an incentive system for fathers to take parental leave. They find that the incentives for male parental leave have a large short-term effect, as males take much more parental leave after the change. However, they do not find behavioral effects in the household such as men having a higher incentive to use their proportion of the leave taken to care for sick children. Cools et al. (2015) show how paternity leave quotas increase the number of fathers taking leave, but also, an improvement in the child's school outcomes. My contribution is presenting evidence on the side effects on households, particularly for the mother. I contribute to the literature by quantifying the effects with a structural model, including a couple selection and intra-household bargaining effects.

The paper is structured as follows: The section that follows provides an overview of the institutional context in Costa Rica, explaining the contextual significance of the Responsible Paternity Law. The third section presents data and empirical evidence on the impact of paternity law on households. The fourth section presents the theoretical model, in which I discuss the effects of selection and intra-household competition. The following section describes the estimation strategy for the structural model. The structural results are presented in section six, and the final section concludes.

2.2 Institutional background

By the year 2000, nearly half of all births in Costa Rica were from single mothers, and one-third of them had no registered father (Robles, 2001). Non-married Costa Rican women had two options for determining the father of their child: first, the man recognized himself as the father or second, they petitioned a judge to order a DNA test. The mother was required to find witnesses and proof of their relationship with the father for the latter. Children born within marriage are not affected by this issue because both parents are automatically registered.

Because of the growing number of children without a registered father, mothers faced the entire cost of raising them. As a result, in 2001, the Costa Rican government proposed and lawmakers passed the Responsible Paternity Law, making it easier for non-married mothers to recognize the father of their newborn child. The paternity law made three important changes: first, even if the presumed father is not present, he can be registered in the hospital[1]. Second, if the man denies being the father, the mother's written and signed statement is sufficient to request a DNA test. The man pays for the test if he is found to be the father or the mother if he is not. Third, the mother receives retroactive child support for pregnancy expenses.

Figure 2.1 shows that after the law was passed, the number of child support demands filed in the Costa Rican Family Court increased. The Family Court received a 26% increase in child

[1]Every hospital in Costa Rica has a Civil Registry office to register births.

support demands in the second quarter of 2001 compared to the same quarter in 2000. This increase undercounts the actual number of child support agreements because it only includes cases where the parents couldn't agree, and the mother had to go to Family Court. The increase in child support demand in 2001 was nearly 16% higher than in 2000. When a request is filed in court, a judge determines a preliminary monetary child support amount until a final agreement is reached.

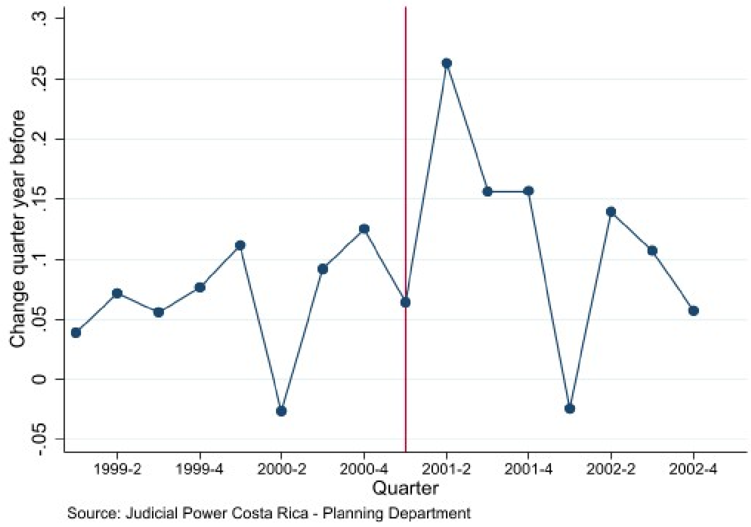


Figure 2.1: **Change in number of child support requests**

There are two studies of the paternity law's effects on women. First, Robles (2001) presents a descriptive analysis of the socioeconomic characteristics of mothers who use the paternity law. She discovers, through interviews with a small sample of women, that women who use the paternity law are mostly non-married women who live with the father of the child, have low education levels, and are mostly unemployed. Second, Ramos-Chaves (2010) shows that the paternity law had a causal effect on fertility outcomes in women. He finds a 5% drop in the birthrate and total fertility rate after the law is implemented. This result is larger for first-time mothers. The decline also had an impact on the marriage rate, implying a drop in marriages due to unexpected pregnancies.

2.3 Data and empirical evidence

2.3.1 Data

I use repeated cross-sections data from 1997 to 2009 of the yearly *Encuesta de Hogares de Propósitos Múltiples* (EHPM) from the Costa Rican *Instituto Nacional de Estadísticas y Censos*

(Instituto Ncional de Estadísticas y Censos, 2009). Every year, the *EHPM* collected approximately 10,000 households and 40,000 individuals. The data includes variables such as age, gender, relationship with the head of the household, marital status, education level, labor situation, monthly wage, hourly working hours, and unemployment.

There is no information in the data to determine whether a woman used the paternity law at the birth of her child. As a result, I must approximate the households that are affected. I apply the same approach as Ramos-Chaves (2010) and set the law's effect effective date to 2002. This is due to the short time gap between the enactment of the bill (April 2001) and data collection (June 2001).

My sample consists of 33,618 households. Each household unit consists of a single individual, a married or cohabiting couple, and other members such as children, parents, or others. I select households where the head woman was at most 33 years old, and the head man was at most 40 years old. The age of the women was chosen based on Ramos-Chaves (2010), which shows that the paternity law has no effect on fertility outcomes for women over the age of 33[2].

Figure 2.2 shows the proportion of men and women working in the subsample. It shows how steady men's labor participation has been between 2001 and 2009, while female participation has almost doubled. Coupled women participate at a lower rate than single women, although their participation is increasing. In terms of weekly working hours, Figure 2.3 reveals another gender gap between men and women, but the average working hours for both groups are rather consistent. Women's behavior varies according to marital status, with married women working fewer hours than single women. For men, the opposite is true; non-married men work less than married men.

[2]Other selections were made, which are explained in Appendix 2.A.

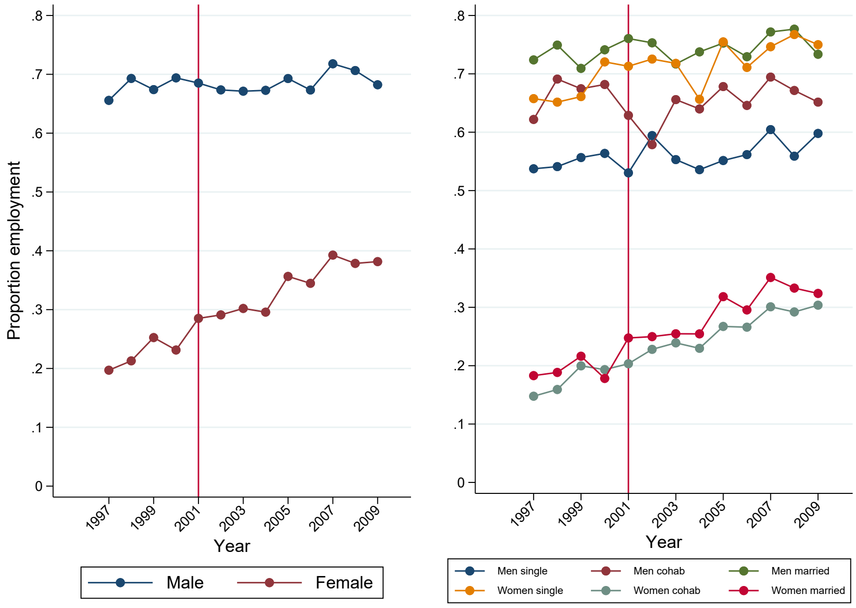


Figure 2.2: Labor Participation

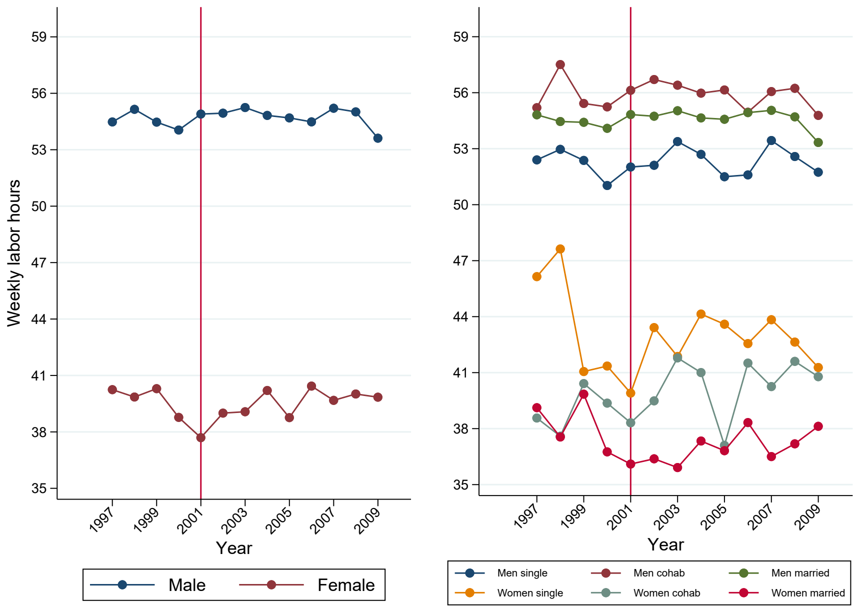


Figure 2.3: Labor Hours

Table 2.1 displays summary statistics for the most important individual variables. I define male employment as full-time employment and unemployment[3]. Female employment considers both full employment and subemployment. The table shows that under these definitions, there

[3]Male labor participation in Costa Rica is extremely high, around 95%. To have more variation I merge unemployment and subemployment. Subemployment is defined as individuals who work more or less than what they would like to.

is a clear gender difference in labor participation and weekly labor hours.

Table 2.1: Individual variables' descriptive statistics

	Men					Women				
Variable	**Obs**	**Mean**	**Std. Dev.**	**Min**	**Max**	**Obs**	**Mean**	**Std. Dev.**	**Min**	**Max**
Single	33,873	0.16		0	1	32,356	0.12		0	1
Cohabit	33,873	0.34		0	1	32,356	0.35		0	1
Married	33,873	0.50		0	1	32,356	0.53		0	1
Age	33,873	34.53	8.85	18.00	64.00	32,356	28.38	5.19	18.00	41.00
Primary School diploma or less	33,873	0.61		0	1	32,356	0.59		0	1
High School diploma or more	33,873	0.39		0	1	32,356	0.41		0	1
Years of Schooling	33,873	7.13	3.59	0.00	19.00	32,356	7.28	3.36	0.00	19.00
Employed	33,873	0.69		0	1	32,356	0.32	0.47	0	1
Hourly wage*	23,107	75.67	43.00	7.50	374.03	10,311	48.08	38.16	3.22	374.17
Labor hours*	23,107	54.69	10.95	14.00	95.00	10,311	39.59	16.81	4.00	80.00

*: Conditional on employment.

Table 2.2 presents the summary statistics for the household variables. Most of the households in my sample live in a rural area outside of the Central Valley. They have two children on average, and the average number of members in the household is 3.6, which can include family members other than the nuclear family. After 2002, 28% had at least one child.

Table 2.2: Summary Statistics - Household Variables

Variable	**Obs**	**Mean**	**Std. Dev.**	**Min**	**Max**
Total household income	37,813	79.32	60.03	0.00	376.52
Total household members	37,813	3.64	1.52	1.00	8.00
Number children					
Zero	37,813	0.23		0	1
One	37,813	0.25		0	1
Two	37,813	0.29		0	1
Three or more	37,813	0.24		0	1
Children born before 2002	37,813	0.66		0	1
Children born after 2002	37,813	0.28		0	1
Geographic urban area					
Outside Central Valley, rural area	37,813	0.44		0	1
Central Valley, rural area	37,813	0.23		0	1
Outside Central Valley, urban area	37,813	0.15		0	1
Central Valley, urban area	37,813	0.18		0	1

2.3.2 Empirical evidence

The law was intended for a non-married woman who is unable to easily declare the father of their newborn child. I consider a treated observation as a woman with a child born after 2002. The control group consists of married women while the treated group groups cohabited and single women. To estimate the paternity law effect, I use the estimation strategy presented by De Chaisemartin and d'Haultfoeuille (2018). Their fuzzy differences-in-differences framework estimates a treatment effect when the proportion of treated units in the treatment and control groups increases, and no unit remains completely untreated.

Consider a binary treatment D. $Y(1)$ and $Y(0)$ represent the potential outcomes of the same treatment unit with and without treatment. The observed outcome is $Y = DY(1)+(1-D)Y(0)$. Consider T a random variable that divides the data into two time periods, before and after 2002. The treated group, $G = 1$, is defined as non-married individuals, which includes cohabited and single individuals. The "sharp" differences-in-differences setting is when $D = G \times T$. In my fuzzy framework, some units in the control group are treated at period 1 while others remain untreated, implying that equality does not hold. This problem arises in my context for two reasons. First, I have repeated cross-sections and do not observe when a household was formed, only when their children were born. Second, married couples could have married after the law was passed, and some control group households did not have a child at the time they were surveyed. Lastly, let $S = \{D(0) < D(1), G = 1\}$ be the set of "treatment group switchers": treatment group units going from non-treatment to treatment between period 0 and period 1. I estimated the Local Average Treatment Effect (LATE) for this group.

The following assumptions must be made to identify the LATE estimator. First, due to the fuzzy design, the control and treatment groups experience a treatment effect, but the latter has a larger increase in the effect of its treatment rate. The second assumption states that the percentage of treated units in the control group remains constant between periods. The third assumption is about treatment status: units are either untreated or treated. Finally, the fourth assumption is a common trend assumption.

Define for any random variable R, R_{gt} and R_{dgt} as two other random variables such that $R_{gt} \sim R|G = g, T = t$ and $R_{dgt} \sim R|D = d, G = g, T = t$, where $\sim$ denotes equality in distribution. Under this notation and the previous assumptions, the LATE estimator is identified and defined as

$$LATE = \frac{E(Y_{11}) - E(Y_{10} + \delta_{D_{10}})}{E(D_{11}) - E(D_{10})}$$

where $\delta_d = E(Y_{d01}) - E(Y_{d00})$ denotes the change in the mean outcome between period 0 and 1 for the control group unit with treatment status d. This estimator is called the time-corrected

Wald estimator (De Chaisemartin and d'Haultfoeuille, 2018).

For dummy outcome variables, the time-corrected Wald estimator works well. De Chaisemartin and d'Haultfoeuille (2018) propose the changes-in-changes Wald ratio for continuous variables. For the latter, the LATE estimator is identified using two additional assumptions. First, the potential outcomes are assumed to be strictly increasing functions for a scalar unobserved heterogeneity term with a stationary distribution over time. Second, there is full support and continuous density of outcomes across all the G and T cells. The LATE is defined as

$$LATE = \frac{E(Y_{11}) - E(Q_{D_{10}}(Y_{10}))}{E(D_{11}) - E(D_{10})}$$

where $Q_d(y) = F^{-1}_{Y_{d01}} \circ F_{Y_{d00}}(y)$ is the quantile-quantile transform of Y from period 0 to 1 in the control group conditional on $D = d$ and F is the density function.

I estimate using the *fuzzydid* Stata package by De Chaisemartin et al. (2019). I estimate the effect of having a child after 2002 on labor participation and weekly hours worked separately for men and women. The results are shown in Table 2.3. Female labor participation is unaffected, but male labor participation is reduced by 8 percentage points. In terms of weekly labor supply, both women and men reduce their supply by 5.6 hours and 4.5 hours, respectively. Some graphs showing the parallel trend assumption are presented in Appendix 2.B.

Table 2.3: **Effect of paternity law in female labor decisions**

	Labor participation		Weekly labor hours	
	Women	Men	Women	Men
LATE	0.03	-0.08**	-5.57*	-4.49**
	(0.045)	(0.037)	(2.935)	(2.040)
Controls	Yes	Yes	No	No
N	32,356	33,694	10,311	23,107

Standard errors computed with a bootstrap procedure using 150 replications. Controls include individual and household demographics and geographical variables.
*:10% significance, **: 5% significance, ***: 1% significance.

These findings indicate that the paternity law had a significant impact on the labor decisions of households. However, the law may have an impact on couple formation. Because men are almost certainly declared fathers under paternity law, they must almost certainly pay child support, which raises the cost of being single. Therefore, given economies of scale in the household, it increases the incentives to form a couple, either married or cohabiting. Women, on the other hand, do not need to marry to receive financial support for their newborn child, and in some cases prefer to be single mothers. This latter effect is consistent with Ramos-Chaves (2010), who found a 5% decrease in the marriage rates. To understand the mechanisms behind the effects of the law on labor participation and labor supply, I present a theoretical model in the following section.

2.4 Theoretical model

In this section, I present a collective marriage matching model following Choo and Seitz (2013). It seeks to explain the couple selection and intra-household effects that paternity law has on economic behavior.

Individuals in the model simultaneously decide whether to form a household and then decide on intra-household allocations. I explain it in two stages for clarity. Individuals decide what type of household they want to form in the first stage. Individuals in my model can choose between single, cohabited, and married households. Wages and assets are known before deciding whether to form a couple, and the bargaining power of the woman and man is determined alongside the option to form a couple. In the second stage, intra-household allocations are chosen to realize the indirect utilities predicted in the first stage. Labor decisions in the household differ between the woman, who chooses her labor supply, and the man, who decides whether to participate in the labor market. This is due to the data's small variation in men's labor supply.

2.4.1 Preferences

Let C_i be private consumption for the man or woman $(i = m, f)$, h_i is i's labor supply and k is the type of household: single (s), cohabited (c) or married (j). Each individual utility is:

$$u(i,k)(1-h_i, C_i) + \Gamma_{i,k} + \epsilon_{i,k},\ i = m, f; \quad k = s, c, j;$$

where the first term is defined over consumption and leisure and affects the intra-household allocation. $\Gamma_{i,k}$ captures invariant gains of being in a household of type k and it is assumed to be separable from consumption and leisure. Lastly, $\epsilon_{i,k}$ is an idiosyncratic, additive separable and i.i.d. preference shock specific to each individual and type of household. The shocks are realized before the marriage decision is made. Both $\Gamma_{i,k}$ and $\epsilon_{i,k}$ affect marriage behavior but do not directly influence the intra-household allocation.

2.4.2 Intra-household allocation

I present the model recursively, starting with the intra-household decision process. It follows the collective model developed by Choo and Seitz (2013) and Blundell et al. (2007). I first describe the single individual household and then the allocation problem for cohabited and married households.

Singles

A single individual faces the problem:

$$\max_{h_i, C_i} \; u(i,s)(1-h_i, C_i) + \Gamma_{i,s} + \epsilon_{i,s}, \;\; i = m, f \tag{2.1}$$

$$\text{s.t.} \;\; C_i = w_i h_i + y_s \quad \text{for } i = m, f \tag{2.1a}$$

where w_i is the wage and y_s is non-labor income when single. This non-labor income includes monetary child support received by the mother and paid for by the father.

Couples

When a man and woman decide to form a couple, whether married or cohabit, they engage in a bargaining process to allocate the household's income and consumption. The following maximization problem defines the collective model:

$$\begin{aligned} \max_{h_m, C_m, h_f, C_f} \;\; & \lambda_k(w_m, w_f, y, \mathbf{z}) u(m,k)(1-h_m, C_m) + \Gamma_{m,k} + \epsilon_{m,k} + \\ & (1-\lambda_k(w_m, w_f, y, \mathbf{z})) u(f,k)(h_f, C_f) + \Gamma_{f,k} + \epsilon_{f,k}, \;\; k = c, j \end{aligned} \tag{2.2}$$

$$s.t. \begin{cases} u(f,k)(1-h_f, C_f) + \Gamma_{f,k} + \epsilon_{f,k} \geq U_s^f(1-h_f, C_f) + \Gamma_{f,s} + \epsilon_{f,s}\,, \;\; k = c, j & \text{(2.2a)} \\ u(m,k)(1-h_m, C_m) + \Gamma_{m,k} + \epsilon_{m,k} \geq U_s^m(1-h_m, C_m) + \Gamma_{m,s} + \epsilon_{m,s}\,, \;\; k = c, j & \text{(2.2b)} \\ C_m + C_f = w_m h_m + w_f h_f + y_k\,, \;\; k = c, j & \text{(2.2c)} \\ h_m \in \{0,1\}, \; 0 \leq h_f \leq 1 & \text{(2.2d)} \end{cases}$$

where $\lambda(w_m, w_f, y, \mathbf{z})$ is a Pareto weight that depends on the couple's wage, non-labor income and distribution factors $\mathbf{z}$ which are defined ahead.

The decision process, as is commonly assumed in collective models, results in Pareto-efficient outcomes[4]. As a result, the model can be decentralized by the Second Welfare Theorem and described as a two-stage process. In the absence of public goods, the Pareto weight has a one-to-one relationship with the bargaining function.

The man and woman allocate total income shares $\Psi_k^m(w_f, w_m, y, \mathbf{z})$ and $\Psi_k^f(w_f, w_m, y, \mathbf{z})$ in the first stage, referred to in the literature as the sharing rules or bargaining functions. These shares are determined by wages, non-labor income, and distribution factors that account for the power each spouse has in the household. This power is based on the man's and woman's outside options: the greater their outside option, the more likely they are to split the couple and be single. This power is linked to a key component of collective models: distribution factors $\mathbf{z}$. A distribution factor is a variable that meets two criteria: (i) it has no effect on preferences or budget constraints, but (ii) it can influence the decision process by influencing the decision power of household members. Some common distributionns factors are the gender ratio in the household's neighborhood and divorce laws (Chiappori et al., 2002). The central idea is that a distribution factor influences a spouse's outside option benefit, increasing his or her bargaining power and improving utility. In the short run, the paternity law is a distribution factor[5].

In the second stage, the man and woman solve their individual problem using income share from the income allocation derived from the first stage. I present the general program faced by the woman given the labor participation decision made by the man:

$$\max_{h_f, C_f} \; u(f,k)(h_f, C_f), \quad k = c, j \tag{2.3}$$

$$s.t.\begin{cases} C_f = w_f h_f + \Psi_k^f(w_f, w_m, y_k, \mathbf{z}), \quad k = c, j & \text{(2.3a)} \\ 0 \leq h_f \leq 1 & \text{(2.3b)} \end{cases}$$

where $\Psi_k^f(w_f, w_m, y_k, \mathbf{z}) = y_k + w_m - \Psi_k^m(w_f, w_m, y_k, \mathbf{z})$ is woman's sharing rule. The Marshallian labor supply of the programme is $H^f(w_f, \Psi_k^f(w_f, w_m, y_k, \mathbf{z}))$ and the reduced form equation is

$$h^f(w_f, w_m, y_k, \mathbf{z}) = H^f[w_f, \Psi_k^f(w_f, w_m, y_k, \mathbf{z})]. \tag{2.4}$$

The man's labor participation frontier is defined by a set of wages and non-labor income bundles (w_f, w_m, y_k) for which the man is indifferent between participating or not. Following Blundell et al. (2007), it is possible to parametrize the labor participation frontier with the use

[4] A simple argument in favor of this assumption is that man and woman can know well each other's preferences and because of their interaction are unlikely to not consider Pareto-improving decisions. For more about the validation of Pareto-efficiency, see Vermeulen (2002) and Chiappori and Mazzocco (2017).

[5] In the long run, the law can potentially affect preferences. I do not consider this case as there is no empirical way to prove it due to data limitations.

of a shadow wage condition and recover the structural parameters of the bargaining function. I explain in section five the identification of the structural parameters and how to retrieve them from a reduced form estimation.

Intuitively, the collective bargain is reflected in both the man's labor participation and the woman's labor supply. Changes in a household's income include the bargaining effect. Because the man's wage increases the man's bargaining power, there is a difference in the woman's labor supply depending on whether extra income comes from non-labor income or an increase in the man's wage. There is also an effect even if the man does not participate in the labor market and does not have a wage because his potential wage gives him bargaining power. The bargaining effect is reflected in the reservation wage that defines the man's labor participation. Changes in the woman's wage have an impact on the reservation wage due to changes in the bargaining process, and thus his share of resources.

2.4.3 Marriage decision and marriage market

In the first stage of the model, once the idiosyncratic gains from couple formation $\epsilon_{i,k}$ are realized, the man and woman decide whether to form or not a couple and what type of couple, either married or cohabited. For individual i, the indirect utility functions for being single, cohabited, or married are respectively:

$$V_{i,s}(\epsilon_{i,s}) = Q_{i,s}[w_i^*, y_s] + \Gamma_{i,s} + \eta_{i,s} \tag{2.5}$$

$$V_{i,c}(\epsilon_{i,c}) = Q_{i,c}[\Psi_c^i(w_f^*, w_m^*, y_c, \mathbf{z})] + \Gamma_{i,c} + \eta_{i,c} \tag{2.6}$$

$$V_{i,j}(\epsilon_{i,j}) = Q_{i,j}[\Psi_j^i(w_f^*, w_m^*, y_j, \mathbf{z})] + \Gamma_{i,j} + \eta_{i,j} \tag{2.7}$$

where $Q_{i,k}[\cdot]$ are the indirect utilities from the second stage intra-household allocation decisions and w_i^* denotes potential wages. The optimal choice is such that

$$\max V_i = \max[V_{i,s}, V_{i,c}, V_{i,j}] \tag{2.8}$$

Under the assumption that the idiosyncratic shocks $\eta_{i,k}$ are i.i.d Type 1 Extreme Value and $V_{i,s}^*$ the indirect utility functions without $\eta_{i,k}$, I can define π_i the probability i prefers to enter a household type k relative to the other alternatives:

$$\pi_{i,k} = \frac{\exp(V_{i,k}^*)}{\sum_{l \in s,c,j} \exp(V_{i,l}^*)} \tag{2.9}$$

The equilibrium definition and proof of existence can be found in Choo and Seitz (2013).

2.4.4 Effect of the Responsible Paternity Law

There are two effects of the paternity law on economic behavior. A couple formation effect and an intra-household effect. First, because the law provides monetary child support to the non-married woman in case of a pregnancy[6], it increases the woman's income, relaxing her budget constraint (equation 2.1). With it, the indirect utility of remaining single (equation 2.5) in a potential pregnancy increases and hence the probability for the woman to be single $\pi_{f,s}$.

For the man, the effect is the opposite: they must pay child support, reducing their available income and, as a result, decreasing their utility of being single. He would try to increase the woman's indirect utility of being in a couple, whether married or cohabited because living in a household implies sharing costs and other benefits (household work, joint income, and so on). The only way he can do that is by increasing the woman's bargaining power, $\Psi_f(\cdot)$, and reducing his own, $\Psi_m(\cdot)$. This increase in the woman's bargaining power equates to an increase in her income, relaxing her budget constraint in equation 2.3a, creating an income effect on her labor supply. This increases the indirect utilities, 2.6 and 2.7, and thus the probability of being in a couple, either cohabit $\pi_{m,c}$ or married $\pi_{m,j}$.

The woman's income effect increases her likelihood of choosing a relationship over being single, which is what the man desired. The final decision is made based on the value $\Gamma_{i,k}$ to determine the type of household.

2.5 Empirical model

In my data, I cannot observe who is the father of the baby for non-married single mothers or the amount of child support for each child. As a result, I am unable to fully estimate the model proposed by Choo and Seitz (2013). However, I conduct two separate estimations. First, I compute a multinomial logit model to quantify the couple formation effect. Second, I estimate the collective household decision-making by Blundell et al. (2007).

2.5.1 Couple formation

As I have repeated cross-sections, the notation i, t denotes an observed individual i at period t. To estimate the couple formation effect, I used a multinomial logit estimation.

Each individual i decides on three possible households k: single (s), cohabited (c) or married

[6]There may be a dynamic effect for married women because they can divorce and have a child later in another relationship and benefit from the law, increasing their outside options. This would increase divorce rates and the likelihood of being single. I ignore it because it is impossible to estimate it using my data.

(j). The utility that individual i obtains from alternative k is decomposed into (1) an observed part labeled $V_{i,k}$ and (2) $\varepsilon_{i,k}$ is an i.i.d random variable. The probability that i chooses alternative k is:

$$\begin{aligned} P_{i,k} &= \text{Prob}(V^*_{i,k} + \varepsilon_{i,k} > V^*_{i,k'} + \varepsilon_{i,k'}, \;\; \forall k' \neq k) \\ &= \text{Prob}(V^*_{i,k} + \varepsilon_{i,k} - V^*_{i,k'} > \varepsilon_{i,k'}, \;\; \forall k' \neq k) \end{aligned}$$

Assuming $\varepsilon_{i,k}$ follows an extreme type 1 value distribution, the probability individual i chooses option k is:

$$P_{i,k} = \frac{e^{V^*_{i,k}}}{\sum_l e^{V^*_{i,l}}}$$

I include in $V^*_{i,k}$ the age, gender, and education of each individual. I control for geographical variables, household information regarding the number of members in the house, the number of children born before 2002, and wealth variables.

2.5.2 Collective household model

The following notation i, t denotes an observed household at period t. The woman's labor supply equation differs depending on the man's labor participation:

$$h^f_{i,t} = A^f_{0,t} + A_m \text{ln} w^m_{i,t} + A_f \text{ln} w^f_{i,t} + A_y y_{i,t} + \mathbf{A} \cdot \mathbf{X}'_{i,t} + u_{1,i,t}, \;\; \text{if husband works} \tag{2.10}$$

$$h^f_{i,t} = a^f_{0,t} + a_m \text{ln} w^m_{i,t} + a_f \text{ln} w^f_{i,t} + a_y y_{i,t} + \mathbf{a} \cdot \mathbf{X}'_{i,t} + u_{0,i,t} \;\; \text{if husband does not work} \tag{2.11}$$

where $\mathbf{X}$ is a vector of control variables that includes the spouses' age and education, geographic and household variables. The man's latent labor participation is:

$$p^m_{i,t} = b^m_{p,t} + b^m_m w^m_{i,t} + b^m_f \text{ln} w^f_{i,t} + b^m_y y_{i,t} + \mathbf{b} \cdot \mathbf{X}'_{i,t} + u^m_{p,i,t} \tag{2.12}$$

I model wages using a standard human capital approach with time variation in the coefficients:

$$w^m_{i,t} = \alpha^m_{0,t} + \alpha^m_{1,t} \text{educ}^m_{i,t} + \alpha^m_{2,t} \text{age}^m_{i,t} + \alpha^m_{3,t} (\text{age}^m_{i,t})^2 + \mathbf{c} \cdot \mathbf{W}'_{i,t} + u^m_{w,i,t} \tag{2.13}$$

$$\text{ln} w^f_{i,t} = \alpha^f_{0,t} + \alpha^f_{1,t} \text{educ}^f_{i,t} + \alpha^f_{2,t} \text{age}^f_{i,t} + \alpha^f_{3,t} (\text{age}^f_{i,t})^2 + \mathbf{d} \cdot \mathbf{W}'_{i,t} + u^f_{w,i,t} \tag{2.14}$$

I assume that the individual's wage is determined solely by her or his age and education, as opposed to the labor outcomes, which are determined by both spouses. Variables in $\mathbf{W}$ only affect wages, as the firm's size, public or private employment, and job position.

Non-labor income is calculated as the difference between the total household income and the total labor income of the spouses. Doing this reduces measurement error and accounts for wealth that may have been overlooked when declaring. Following Blundell et al. (2007), I regard this measure as endogenous and use its predicted values using the reduced form equation:

$$
\begin{aligned}
y_{i,t} = \alpha^y_{0,t} + \alpha^y_{1,t}\text{educ}^m_{i,t} + \alpha^y_{2,t}\text{age}^m_{i,t} + \alpha^y_{3,t}(\text{age}^m_{i,t})^2 + \alpha^y_{4,t}\text{educ}^f_{i,t} + \alpha^y_{5,t}\text{age}^f_{i,t} + \\
\alpha^y_{6,t}(\text{age}^f_{i,t})^2 + \alpha^y_{7,t}\mu_{i,t} + \alpha^y_{8,t}\mu^2_{i,t} + \alpha^y_{9,t}\mathbf{1}(\mu_{i,t} > 0) + \mathbf{e} \cdot \mathbf{K}'_{i,t} + u_{y,i,t}
\end{aligned} \tag{2.15}
$$

where $\mathbf{K}_{i,t}$ are household variables that include geographical location and the total number of members. $\mu_{i,t}$ includes non-labor income transfers, and other household members' wages and transfers. I include a quadratic term and a nonzero indicator to improve the fit.

Identification

The identification of bargaining parameters is based on *double indifference* used by Blundell et al. (2007). This assumption states that if member i is indifferent about whether to work, then the other member is indifferent too. This assumption implies that the model's solution is contingent on whether the man participates. However, the sharing function is continuous in the participation frontier, allowing for identification. The model's constraints are detailed in Appendix 2.D. It includes the entire model as well as instructions on how to recover structural parameters from reduced-form results.

2.6 Results

First, I present the couple formation model estimated using multinomial logit. The average marginal effects on the probability of each marital status category are shown in Table 2.4. The man is less likely to be single if he had a child after the paternity law was passed. The likelihood that he is in a couple, whether cohabit or married, rises. The woman, on the other hand, is more likely to be a single mother or cohabit, and less likely to be married. These findings corroborate the model's predictions, showing a couple selection effects from the law.

Table 2.4: Average Marginal Effects of paternity law on Marital Status

	All sample	Men	Women
Single	0.01**	-0.16***	0.07***
	(0.004)	(0.011)	(0.005)
Cohabit	0.05***	0.11***	0.03***
	(0.006)	(0.007)	(0.006)
Married	-0.06***	0.05***	-0.10***
	(0.006)	(0.009)	(0.006)
Controls	Yes	Yes	Yes
N	66,229	33,873	32,356

Standard errors under parentheses clustered at the household year level.
Controls include individual and household demographics and geographical variables.
*:10% significance, **: 5% significance, ***: 1% significance.

Next, I present the findings from the estimation of the collective household model. I use the couples' sample where the woman works as I only recover the bargaining function if $h^f_{i,t} > 0$ (Blundell et al., 2007).

To estimate the bargaining effect of the paternity law on households, I estimate the collective household model on two different samples based on whether the paternity law affected the households. I define treated as those cohabited households who had a child after 2002, while cohabited households without a child born children after 2002 serve as the control sample. The main result is determining whether the man and woman collectively bargain in the household or not, and the intra-household effect of the paternity law. For each sample, I estimate two models: an unrestricted model and a model with collective restrictions, which are both reported in the appendix. The likelihood ratio test is whether the collective constraints hold. The null hypothesis for the collective model is that in the household, both the man and woman in the cohabiting household are decision-makers. As mentioned above, households in a collective framework react differently to non-labor income and wages, as the bringer of them gains bargaining power. The alternative hypothesis is not clear as it is difficult to tell from which channel the rejection of the null hypothesis comes. However, if the household does not behave as in the collective model, it might do as in the unitary model, where there is a single decision-maker. If this is the case, income is pooled, such that an increase in non-labor income or in the man's wage has the same effect on the woman's labor supply (only if the man works). The extra income has a standard income effect on the man's labor participation. I present the theoretical unitary model and results in the appendix.

The results for woman's labor supply and man's participation in the control sample is presented in Table 2.5 and the result for the treated sample is in Table 2.6. The complete tables results are presented in Appendix 2.F.

Table 2.5: Estimation results - treated sample

Woman's weekly labor hours - Man does not work				
	Unrestricted		Collective	
	Coef	**SE**	**Coef**	**SE**
Hourly wage man	-3.667	(2.430)	11.484	(0.337)
Hourly wage woman	2.114	(4.472)	10.387	(0.415)
Non-labor income	-4.688	(9.764)	-6.290	(0.031)
Intercept	35.762	(8.606)	6.254	(0.135)
N	1,376		1,376	

Woman's weekly labor hours - Man works				
	Unrestricted		Collective	
	Coef	**SE**	**Coef**	**SE**
Hourly wage man	-1.405	(2.430)	-2.641	$(3.364)^{\dagger}$
Hourly wage woman	4.965	(2.184)	9.341	$(2.172)^{\dagger}$
Non-labor income	0.113	(1.180)	-6.284	(0.030)
Intercept	39.684	(5.061)	15.127	(0.450)
N	2,622		2,622	

Man's labor participation				
	Unrestricted		Collective	
	Coef	**SE**	**Coef**	**SE**
Hourly wage man	1.037	(0.181)	0.833	(0.121)
Hourly wage woman	-0.060	(0.157)	0.062	(0.133)
Non-labor income	-0.343	(0.332)	0.000	(0.003)
Intercept	-0.731	(0.262)	-0.633	(0.238)
N	3,998		3,998	

Standard errors are under parenthesis and have been computed using the bootstrap with 1000 repetitions and allowing for the fact that man's wage, woman's wage and non-labor income are predicted.
All estimations control for geographical zone, number of children, age, high school or college diploma, and year effects.
†: *standard errors computed using the Delta Method.*

Table 2.6: Estimation results - treated sample

Woman's weekly labor hours - Man does not work				
	Unrestricted		Collective	
	Coef	**SE**	**Coef**	**SE**
Hourly wage man	6.557	(4.106)	9.633	(1.204)
Hourly wage woman	14.584	(5.490)	7.506	(1.681)
Non-labor income	-12.408	(15.343)	-21.294	(1.299)
Intercept	31.105	(12.096)	23.304	(3.837)
N	1,506		1,506	

Woman's weekly labor hours - Man works				
	Unrestricted		Collective	
	Coef	**SE**	**Coef**	**SE**
Hourly wage man	-0.047	(4.106)	-0.398	(2.941)†
Hourly wage woman	8.937	(3.463)	10.365	(2.194)†
Non-labor income	-22.812	(0.747)	-21.472	(2.303)
Intercept	45.158	(6.366)	47.059	(4.592)
N	2,868		2,868	

Man's labor participation				
	Unrestricted		Collective	
	Coef	**SE**	**Coef**	**SE**
Hourly wage man	0.632	(0.232)	0.640	(0.178)
Hourly wage woman	-0.237	(0.191)	-0.182	(0.138)
Non-labor income	-0.022	(0.433)	0.011	(0.223)
Intercept	-0.906	(0.475)	-0.920	(0.454)
N	4,374		4,374	

Standard errors are under parenthesis and have been computed using the bootstrap with 1000 repetitions and allowing for the fact that man's wage, woman's wage and non-labor income are predicted.
All estimations control for geographical zone, number of children, age, high school or college diploma, and year effects.
†*: standard errors computed using the Delta Method.*

Testing the collective restrictions with the treatment sample, the likelihood ratio statistic is 0.92 with a p-value of 0.63; hence, I do not reject that treated households behave as a collective household. For the control sample, the likelihood ratio statistic is 16.49 with a p-value of 0.0003; hence, I do reject that control households behave in a collective way.

The difference between samples being decision-makers or not is explained from the outside options for the man and woman. A woman in the control sample does not necessarily have child support or any other type of income from the man in case she decides to split the couple. A woman in the treated sample, on the contrary, can have child support from the man, which allows her to bargain from a better position with the man. This difference is reflected in the man's opportunity cost of working. It suggests that the cohabited man with a child born after the paternity law has less negotiating power with respect to the woman, which lowers his income share. The opportunity cost for a cohabited man without a child after the paternity law is part of the decision made by the household. This opportunity cost is in the model via a reservation wage. Recovering the man's reservation wages for both samples:

$$w^r_{treated} = \kappa_{treated} + \underset{(0.285)}{0.212} \ln w_f - \underset{(0.018)}{0.350}\, y, \quad \text{if child after paternity law}$$
$$w^r_{control} = \kappa_{control} + \underset{(0.067)}{0.146} \ln w_f + \underset{(0.371)}{0.326}\, y, \quad \text{if no child after paternity law.}$$

The coefficients show that man's reservation wage in the treated sample decreases for additional non-labor income and increases with the woman's wage, while in the control sample, it increases with both. I cannot test the difference between them as they were estimated using different samples.

Because in the treated sample I do not reject the collective household restrictions, I can recover the bargaining function by mapping the estimated parameters to the structural parameters. The mapping is explained in detail in Appendix 2.D.3. The man's bargaining function from equation 2.6 is[7]:

$$\Psi^m_{treated} = \kappa_1 + \underset{(0.140)}{0.981}\, w_m + \underset{(0.226)}{0.253} \ln w_f - \underset{(0.316)}{0.016}\, y, \quad \text{if man works}$$
$$F(\Psi^m_{treated}) = \kappa_0 + \underset{(0.113)}{0.464}\left(\underset{(0.140)}{0.981}\, w_m + \underset{(0.226)}{0.253} \ln w_f - \underset{(0.316)}{0.016}\, y\right), \quad \text{if man does not work}$$

From the equations, the man receives 0.981 extra units from the household total income for an additional unit in his wage. He obtains -0.016 extra units for every unit of non-labor income in the household. In exchange, the woman transfers him 0.253 units for every 10% rise in her wage. Finally, if the man does not work, he receives 46.4% of the total income he receives if he works.

These results show an important difference that comes from having a child once the paternity law took place. A man and woman in a cohabited household behave differently in their decision process. Those with a child after the law bargain collectively between themselves to allocate income, where the man has an important drop in his personal income if he decides not to work.

[7] Standard errors are estimated using the Delta Method.

On the other side, a man and woman that cohabit without a child born after the law, allocate income as a single agent household, indifferently if the man works or not. Hence, the paternity law had an important bargaining effect on the woman, by allowing her to have her own sharing rule.

Last, I estimated these results for different control samples, including married couples, samples with married and couple households, and couples before 2002. Most of the results reject the collective bargaining hypothesis. This goes in line with the model predictions that the law had a bargaining effect on cohabited couples that had a child, but not on married couples. On the latter, there was only a couple formation effect as shown above.

2.7 Conclusion

In this paper, I provide evidence on how a law that allowed non-married mothers to automatically declare the father of their newborn child affected labor decisions in Costa Rican households. I focus on labor outcomes: weekly worked hours and labor participation. The differences-in-differences estimation shows that the paternity law had significant effects on households' labor decisions. It decreased labor participation for both spouses, mostly for the man, and decreased the woman's labor supply.

I present a theoretical model to explain that the paternity law created two effects: a couple selection effect and an intra-household bargaining effect. I obtained a negative effect on the probability of a woman getting married due to the paternity law, where they are more likely to be single or in a cohabited household. The inverse happens for the man. From the estimation of a collective household model, I obtain that cohabited couples who had a child after the paternity law behave accordingly to a collective model, but not those cohabited couples without a child. The couples with children behave such that the man decreases his consumption in case he does not work, and the woman benefits from the man's wage in case he works. However, it is important to point out that I did not include in the model a potential substitution with domestic work, which can give different results depending on the effect of the law on domestic chores.

My results show how paternity laws that directly benefit children instead of the parents have wide effects on the economic behavior of the household. In this case, the registration of the father at birth generates a couple selection effect and affects decision-making in couple households. Specifically, my paper opens the possibility to think more broadly about how child-related laws affect mothers and can have an impact on the labor force. Further research is needed to understand the bargain changes from paternities laws, including the use of domestic time use data. These variables can help understand the substitution in time allocation between spouses.

Bibliography

Blundell, R., Chiappori, P.-A., Magnac, T., and Meghir, C. (2007). Collective labour supply: Heterogeneity and non-participation. *The Review of Economic Studies*, 74(2):417–445.

Chiappori, P.-A. (1992). Collective labor supply and welfare. *Journal of political Economy*, 100(3):437–467.

Chiappori, P.-A., Fortin, B., and Lacroix, G. (2002). Marriage market, divorce legislation, and household labor supply. *Journal of political Economy*, 110(1):37–72.

Chiappori, P.-A., Iyigun, M., Lafortune, J., and Weiss, Y. (2017). Changing the rules midway: the impact of granting alimony rights on existing and newly formed partnerships. *The Economic Journal*, 127(604):1874–1905.

Chiappori, P.-A. and Mazzocco, M. (2017). Static and intertemporal household decisions. *Journal of Economic Literature*, 55(3):985–1045.

Choo, E. and Seitz, S. (2013). The collective marriage matching model: Identification, estimation, and testing. In *Structural Econometric Models*. Emerald Group Publishing Limited.

Cools, S., Fiva, J. H., and Kirkebøen, L. J. (2015). Causal effects of paternity leave on children and parents. *The Scandinavian Journal of Economics*, 117(3):801–828.

De Chaisemartin, C. and d'Haultfoeuille, X. (2018). Fuzzy differences-in-differences. *The Review of Economic Studies*, 85(2):999–1028.

De Chaisemartin, C., D'Haultfœuille, X., and Guyonvarch, Y. (2019). Fuzzy differences-in-differences with stata. *The Stata Journal*, 19(2):435–458.

Ekberg, J., Eriksson, R., and Friebel, G. (2013). Parental leave—a policy evaluation of the swedish "daddy-month" reform. *Journal of Public Economics*, 97:131–143.

Goussé, M. and Leturcq, M. (2022). More or less unmarried. the impact of legal settings of cohabitation on labour market outcomes. *European Economic Review*, 149:104259.

Instituto Ncional de Estadísticas y Censos (1997-2009). Encuesta de Hogares de Propósitos Múltiples.

Organisation for Economic Co-operation and Development (2012). *Closing the gender gap: Act now*. OECD Publishing Paris.

Ramos-Chaves, A. R. (2010). *Essays on Economic Development in Costa Rica*. PhD thesis, UC Berkeley.

Reynoso, A. (2018). The impact of divorce laws on the equilibrium in the marriage market. *Unpublished manuscript*.

Robles, I. V. (2001). Inscripción de los hijos e hijas de madres solteras y paternidad responsable. estudio exploratorio realizado con mujeres atendidas en el hospital de las mujeres y el hospital san juan de dios.

Rossin-Slater, M. (2017). Signing up new fathers: Do paternity establishment initiatives increase marriage, parental investment, and child well-being? *American Economic Journal: Applied Economics*, 9(2):93–130.

Vermeulen, F. (2002). Collective household models: principles and main results. *Journal of Economic Surveys*, 16(4):533–564.

Wodon, Q. T. and De La Briere, B. (2018). *Unrealized potential: The high cost of gender inequality in earnings*. World Bank.

2.A Data manipulation

The data was constructed for the observational unit to be a household. A household can be a single individual or a couple. In the household there can be children or other family members.

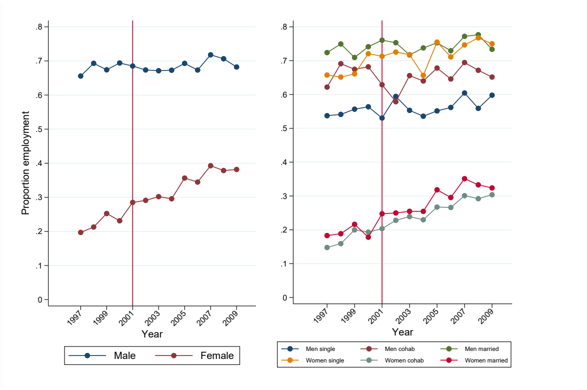
The data cleaning process consisted of dropping a household completely if one of its members presented one of the following characteristics:

- Individuals without age.
- Individuals without years of schooling.
- Households that declared themselves as in a couple but only had one spouse present.
- Households were the age of one or both spouses are over 65.
- Same gender couples.
- Households one spouse was under age 18.
- Households one or both spouses declare themselves as unpaid workers.
- Households one or both spouses declare themselves as working but had a zero wage.
- Households one spouse or both spouses attend education.
- Households one spouse's wage was on the top and bottom 3%.
- Households other income variable was on the top and bottom 2%.
- Households the primary working hours variable was on the top and bottom 1.5%.

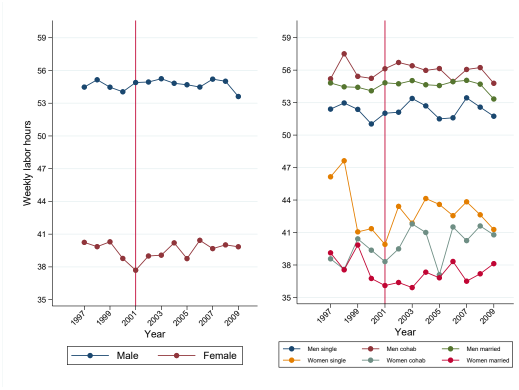
For the wage and income variables, I corrected for inflation using the Costa Rican Central Bank information and setting the base year in July 2009. All the values are in hundred thousand *colones* (exchange rate was 577 *colones* for 1 US dollar).

2.B Fuzzy differences-in-differences

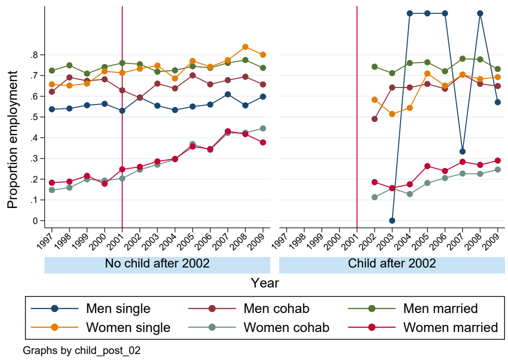
In this section, I present graphs corresponding to the parallel trend assumption for the household. The graphs correspond to the evolution of household formed and labor outputs for men and women, before and after the paternity law.



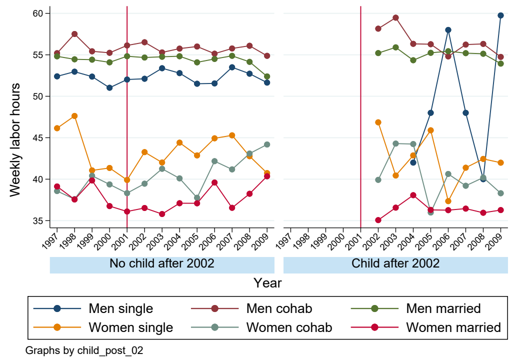
(a) Labor participation



(b) Labor hours



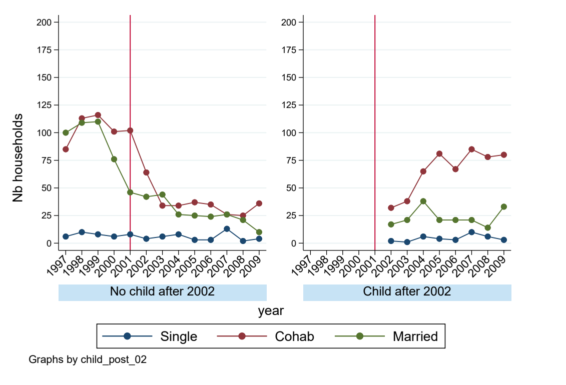
(c) Labor participation by child condition



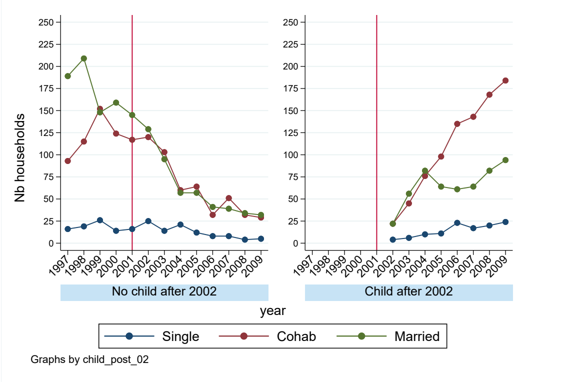
(d) Labor hours by child condition

Figure A1: Labor outputs by marital status

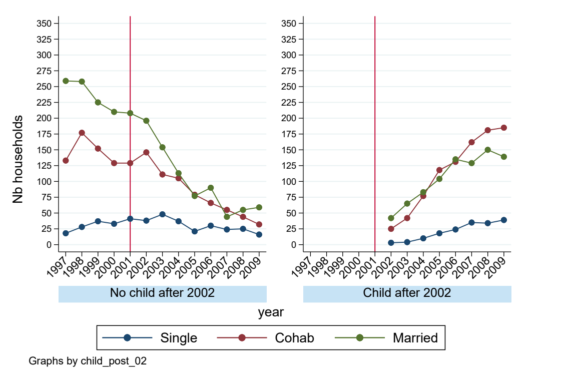
An important characteristic of the treatment relates to women's fertility. Because of this, I also present the graphs according to women's group age.



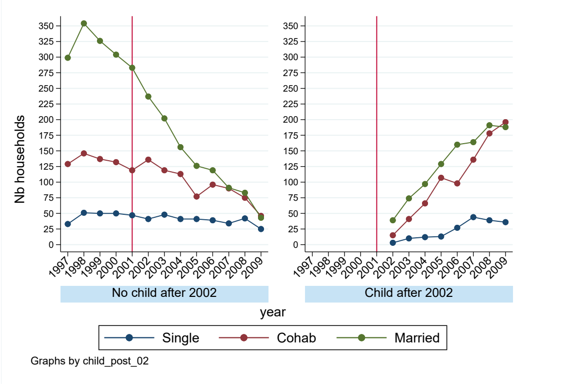
(a) Ages 18-20



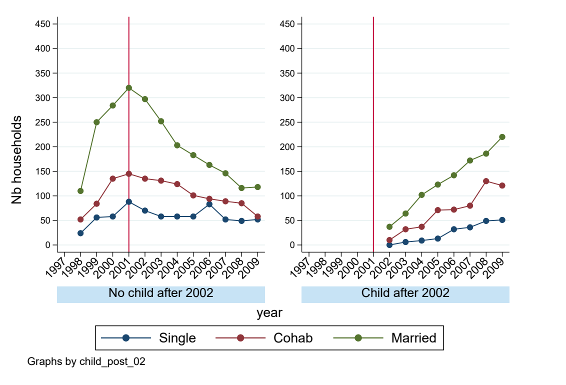
(b) Ages 21-23



(c) Ages 24-26

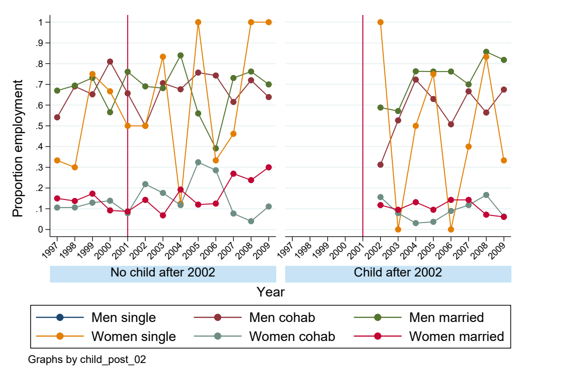


(d) Ages 27-29

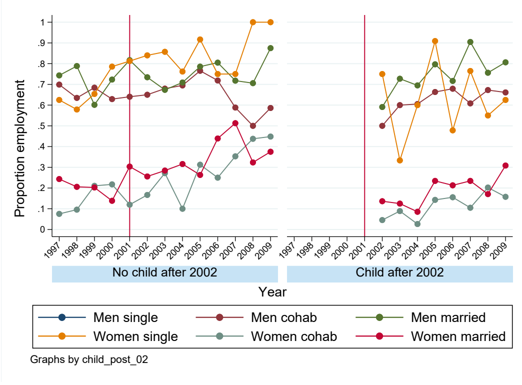


(e) Ages 30-32

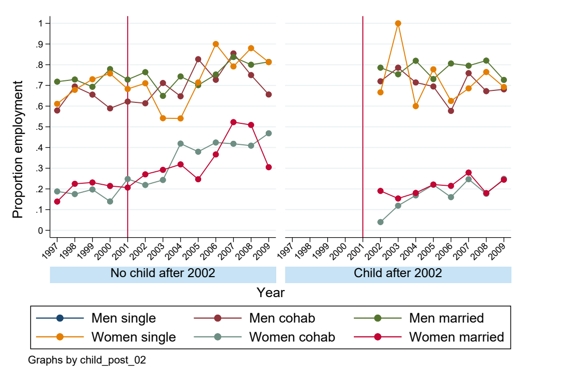
Figure A2: Number of households by martial status and women's age



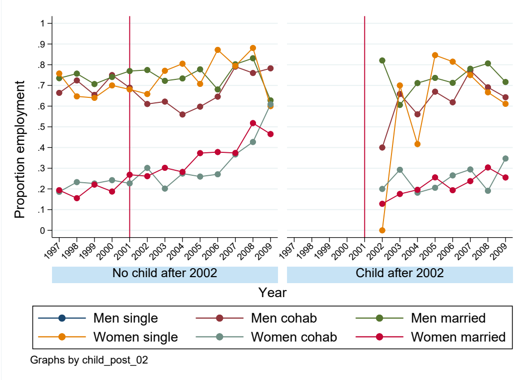
(a) Ages 18-20



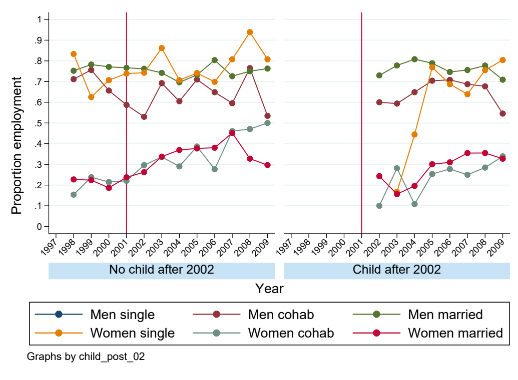
(b) Ages 21-23



(c) Ages 24-26



(d) Ages 27-29

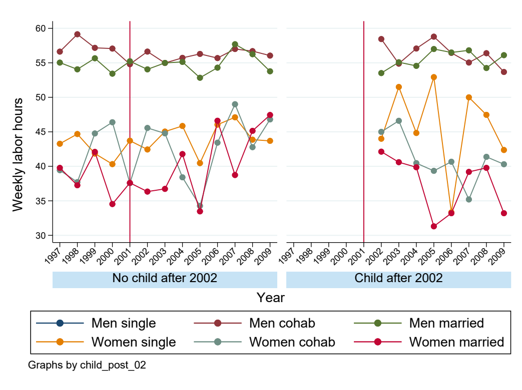


(e) Ages 30-32

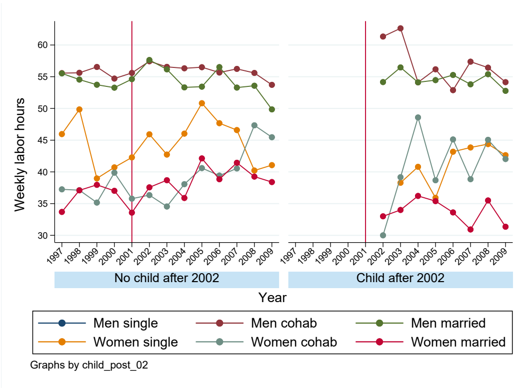
Figure A3: Labor participation by martial status and women's age

(a) Ages 18-20

(b) Ages 21-23



(c) Ages 24-26



(d) Ages 27-29

(e) Ages 30-32

Figure A4: Labor hours by martial status and women's age

2.C Imputation

I impute teens' wages and household non-labor income using different samples of the Costa Rican National Household Survey from 1997 to 2009. I impute men's and women's wages separately using a Heckman two-step selection procedure with the following Mince equation for wages:

$$\log w^i = \alpha_0^i + \alpha_2^p \text{age}^i + \alpha_3^p (\text{age}^i)^2 + \mathbf{X}'\mathbf{B} + u_w^i \tag{A1}$$

where i is an individual and $\mathbf{X}$ includes demographics and exogenous variables related to the industry she works and the size of the firm she is employed. Table A1 shows the results of the estimation. For the women, Table A2 shows the results. Figure A5 shows the comparison between the observed and predicted values for both imputations.

Table A1: Men's wage imputation results

	Dependent variable:	
	Employed	Log Hourly wage rate
Years of education	0.072***	0.057***
	(0.003)	(0.002)
Experience	0.005	
	(0.012)	
Experience square	0.205***	
	(0.030)	
Total number people in hhd	0.301***	
	(0.028)	
Cohabited hhd	0.021	
	(0.032)	
Married hhd	−0.083***	
	(0.031)	
Number of children	−0.060*	
	(0.032)	
Children under age 6		−0.062***
		(0.009)
Children age 7-17		0.068***
		(0.007)
Size firm 1-5		−0.074***
		(0.011)
Size firm 20 or more		−0.075***
		(0.013)
Self-employed		−0.211***
		(0.011)
Employed himself	0.038***	0.037***
	(0.010)	(0.004)
Employed private sector	−0.001***	−0.0005***
	(0.0001)	(0.0001)
Central Valley rural zone	0.149***	0.082***
	(0.021)	(0.008)
Non Central Valley urban zone	0.248***	0.071***
	(0.026)	(0.010)
Central Valley urban zone	0.261***	0.138***
	(0.024)	(0.010)
Constant	−0.761***	−0.891***
	(0.171)	(0.084)
Year effects	Yes	Yes
Observations	25,924	19,229
R^2		0.292
Adjusted R^2		0.291
Log Likelihood	-16,290.130	
Akaike Inf. Crit.	32,630.260	
ρ		0.544
Inverse Mills Ratio		0.220*** (0.056)

*p<0.1, **p<0.05, ***p<0.01.
Baseline categories: different occupations (manager, research, technical and academic professors, and staff), different industries (finance, public administration, real state, teaching, social health, domestic and others), spouse or another relationship in the household, working in a firm with less than 10 employees and for the geographical variable it is living outside the Central Valley in a rural zone.

Table A2: Women's wage imputation results

	Dependent variable:	
	Employed	Log Hourly wage rate
Age	1.274***	0.395***
	(0.102)	(0.073)
Age square	−0.027***	−0.009***
	(0.002)	(0.002)
Head of household	−0.249***	
	(0.090)	
Spouse head of household	−0.349***	
	(0.107)	
Child of household	0.059	
	(0.076)	
Cohabited	−2.483***	
	(0.048)	
Married		−0.164***
		(0.035)
Single		−0.067
		(0.053)
Manual occupation		−0.066
		(0.051)
Size firm 1-5		−0.010
		(0.042)
Size firm 10-19		0.029
		(0.031)
Size firm 100-		−0.204***
		(0.036)
Industry Manufacture	−0.380***	
	(0.088)	
Industry Services	−0.313***	
	(0.076)	
Industry Domestic Services	−0.121	
	(0.083)	
Central Valley rural zone	0.222***	0.104***
	(0.056)	(0.037)
Non Central Valley urban zone	−0.058	0.096***
	(0.054)	(0.036)
Central Valley urban zone	0.150***	0.137***
	(0.047)	(0.031)
Constant	−13.151***	−4.439***
	(1.049)	(0.772)
Year effects	Yes	Yes
Observations	14,272	1,931
R^2		0.115
Adjusted R^2		0.105
Log Likelihood	-2,800.443	
Akaike Inf. Crit.	5,644.886	
ρ		0.148
Inverse Mills Ratio		0.074*** (0.017)

*p<0.1, **p<0.05, ***p<0.01.
Baseline categories: different occupations (manager, research, technical and academic professors, and staff), different industries (finance, public administration, real state, teaching, social health, domestic and others), spouse or another relationship in the household, working in a firm with less than 10 employees and for the geographical variable it is living outside the Central Valley in a rural zone.

For non-labor income, I define it as the difference between household labor income and its total income. The sample of I use consists of 46,568 households. I impute using the predicted

Figure A5: Imputation wages

(a) Women. (b) Men.

values from the following regression:

$$y = \alpha_0^y + \alpha_1^y \text{educ}_i^f + \alpha_2^y \text{age}_i^f + \alpha_3^y (\text{age}^f)_i^2 + \alpha_4^y \text{educ}_i^m + \alpha_5^y \text{age}_i^m + \alpha_6^m (\text{age}^m)_i^2 + \alpha_7^a \mathbf{1}(a_i > 0) + \mathbf{Q}_i' \mathbf{E} + u_{yi} \tag{A2}$$

where i is a household, f refers to the father, m the mother, and a are the rents and profits that the household has, unrelated to labor and government transfers. I include it as an indicator for $a > 0$ to improve the fit. $\mathbf{Q}_i$ are household-level variables like the number of children, demographics, and geographical location. Table A3 shows the results of the estimation and Figure A6 shows the comparison between the observed and predicted values.

Table A3: Non-labor income imputation results

	Dependent variable:
	Non-labor income
Zero rent income	−32.606***
	(1.907)
Age father	0.002
	(0.909)
Age father square	0.001
	(0.010)
Age mother	1.049
	(1.110)
Age mother square	−0.012
	(0.013)
Father's years of schooling	2.306***
	(0.179)
Mother's years of schooling	2.849***
	(0.180)
Household of 4	0.086
	(1.455)
Household of 5	3.101*
	(1.763)
Household of 6	10.128***
	(2.656)
Central Valley rural zone	−4.126***
	(1.579)
Non Central Valley urban zone	−3.109*
	(1.684)
Central Valley urban zone	−4.715***
	(1.511)
Constant	27.228
	(25.126)
Year effects	Yes
Observations	2,756
R^2	0.321
Adjusted R^2	0.316
Residual Std. Error	28.778 (df = 2,734)
F Statistic	61.668*** (df = 21; 2,734)

*p<0.1, **p<0.05, ***p<0.01.

Baseline categories: positive rent income, household members, and for the geographical variable it is living outside the Central Valley in a rural zone.

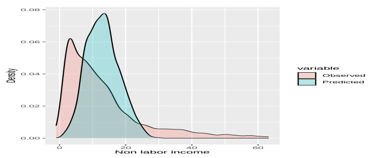


Figure A6: Imputation non-labor income

2.D Collective household model

2.D.1 Extended theoretical model

In this section, I present the extended model presented by Blundell et al. (2007). Their model depends on the husband's labor participation and the wife's labor supply.

Wife's labor decision when the husband works If the husband participates in the labor market his utility is $U_k^m(0, C^m)$, then:

$$u(m,k)(0, C^m) = \bar{u}^m(w_f, w_m, y_k), \quad k = c, j \tag{A3}$$

Solving (A3) for C^m:

$$C^m = \Psi_k(w_f, w_m, y_k)$$

$\Psi_k(w_f, w_m, y_k)$ is the sharing rule, which is affected by wages and non-labor income. With the solution of C^m and because of Pareto efficiency, the wife's optimal decision is the solution of the following programme:

$$\max_{h_f, C_f} \; u(f,k)(1 - h_f, C_f), \quad k = c, j \tag{A4}$$

$$s.t. \begin{cases} C_f = w_f h_f + \Psi_k^f(w_f, w_m, y_k) & \text{(A4a)} \\ 0 \leq h_f \leq 1 & \text{(A4b)} \end{cases}$$

where $\Psi_k^f(w_f, w_m, y_k) = y_k + w_m - \Psi_k^m(w_f, w_m, y_k)$. The solution of the programme is $H^f(w_f, \Psi_k^f(w_f, w_m, y_k))$ and the reduced form is:

$$h^f(w_f, w_m, y_k) = H^f[w_f, \Psi_k^f(w_f, w_m, y_k)] \tag{A5}$$

Wife's labor decision when the husband does not work

In the case the husband does not work, his utility is $U_k^m(1, C^m)$ and:

$$u(m,k)(1,C^m) = \bar{u}_k^m(w_f, w_m, y_k) \tag{A6}$$

which can be solved to:

$$C^m = F(\Psi_k(w_f, w_m, y_k))$$

where $F(\cdot)$ is a transformation to consider the fact that the man does not work. The wife's decision program can be written as above, and it leads to a labor supply of the form:

$$h^f(w_f, w_m, y_k) = H^f[w_f, y_k - F(\Psi_k(w_f, w_m, y_k))] \tag{A7}$$

Husband's labor participation decision

The participation frontier L is defined by a set of wages and non-labor income bundles (w_f, w_m, y_k) for which the husband is indifferent between participating or not. Using *Lemma 1* in Blundell et al. (2007), it is possible to parametrize L with the use of a shadow wage condition,

$$w_m > \gamma(w_f, y)$$

for some γ that describes the participation frontier, which is true with the following assumption from Blundell et al. (2007).

Assumption. *The sharing rules are such that:*

$$\forall(w_f, w_m, y_k), \quad \left|[1 - F'(\Psi_k(w_f, w_m, y))] \times \frac{\partial \Psi_k(w_f, w_m, y_k)}{\partial w_m}\right| < 1 \tag{A8}$$

So whenever $h^f > 0$, γ is characterized by:

$$\forall(w_f, y_k), \quad \Psi_k(w_f, w_m, y_k) - F(\Psi_k(w_f, \gamma(w_f, y), y_k)) = \gamma(w_f, y_k) \tag{A9}$$

2.D.2 Restrictions

To recover the collective model structural parameters from a labor supply reduced form estimation it is necessary to add restrictions. First, from the male participation equation (2.12) solving for w_{it}^m when $p_{i,t}^m = 0$ allows obtaining the male reservation earnings and the parameters on the husband's participation frontier:

$$\gamma_f = -\frac{b_f^m}{bm^m}, \qquad \gamma_{y_k} = -\frac{b_{y_k}^m}{bm^m} \tag{A10}$$

Second, to recover the wife's labor structural parameters the restrictions come from equations (2.10), (2.11) and (A10):

$$-\frac{1}{\gamma_{y_k}} = \frac{A_m - a_m}{A_{y_k} - a_{y_k}}, \qquad \frac{\gamma_f}{\gamma_{y_k}} = \frac{A_f - a_f}{A_{y_k} - a_{y_k}} \tag{A11}$$

2.D.3 Recovering structural parameters

If the data do not reject the collective restrictions, I can recover the sharing function of the household as done by Blundell et al. (2007). On any point of the frontier, the four restrictions above, create a non-linear system of equations in the unknowns $(\psi_f, \psi_m, \psi_y, F')$. With some algebra, one obtains the following equation in F':

$$(\gamma_y ba - 1 + a - \gamma_y b)(F')^2 + (-b + 1 - 2\gamma_y ba + \gamma_y a - a)F' + b + \gamma_y ba = 0$$

where $a = a(w^f, y) = A[w^f, \gamma(w^f, y), y]$ and likewise for b. Blundell et al. (2007) show that if there is a solution to this quadratic equation that satisfies equation (A8), then the sharing rule is identified. This solution is such that:

$$F'(\Psi^m(w^f, w^m, y)) = \theta_\Psi^m(w^f, y)$$

and (ψ_f, ψ_m, ψ_y) are recovered with the following equations (rewritten from the restrictions above):

$$\psi_m[w^f, \gamma(w^f, y), y] = K(w^f, y) = \frac{b}{(a-b)}\left(a - 1 - \frac{a}{\theta_\Psi^m(w^f, y)}\right)$$
$$\psi_f[w^f, \gamma(w^f, y), y] = L(w^f, y) = \frac{\gamma_f}{(a-b)\gamma_y}\left(a - 1 - \frac{b}{\theta_\Psi^m(w^f, y)}\right)$$
$$\psi_y[w^f, \gamma(w^f, y), y] = M(w^f, y) = \frac{1}{(a-b)}\left(a - 1 - \frac{b}{\theta_\Psi^m(w^f, y)}\right)$$

Then, from the mapping between the structural Marshallian labor supply and its reduced form equation, I recover the last parameters:

$$\theta_\Psi^f = \frac{A_y}{1 - \psi_y}$$
$$\theta_w^f = A_f + \theta_\Psi^f \psi_f$$

Lastly, it is important to remind that functions Ψ and F are identified up to a constant on the man's labor participation.

2.D.4 Identification and stochastic specification

The identification of the model comes from various sources. First, the sharing rule when both partners are working is a result of Chiappori et al. (2002). This sharing rule is identified up to an additive constant.

There are two restrictions for each case whether the husband participates in the labor market or not. For any (w_f, w_m, y_k) and $h_f((w_f, w_m, y_k)) > 0$:

$$\begin{aligned} A(w_f, w_m, y_k) &= \frac{h^f_{w_m}}{h^f_{y_k}}, \text{ if husband works} \\ B(w_f, w_m, y_k) &= \frac{h^f_{w_m}}{h^f_{y_k}}, \text{ if husband does not work} \end{aligned} \tag{A12}$$

These restrictions show the fact that, when the man is working, his wage affects the woman's labor supply only through an income effect. In other words, what is determined by the couple's decision process is the man's reserved utility he would reach for each wage-income bundle. This utility's level is implemented by distinct levels of consumption, affected by the husband's labor participation.

The second identification is from Blundell et al. (2007) for collective models with corner solutions. The main assumption is "double indifference". It states that, in the participation frontier, both spouses are indifferent between one spouse working or not:

$$\begin{aligned} (\Psi_{y_k} &+ \gamma_{y_k}\Psi_{w_m}) = \frac{\gamma_y}{1-F'} \\ \Psi_{w_m} &= \frac{\gamma_{w_f}}{\gamma_{y_k}}\Psi_{y_k} \end{aligned} \tag{A13}$$

Restrictions (A12) and (A13) create a system of partial derivatives for $\Psi_{w_f}, \Psi_{w_m}, \Psi_y, F'(\cdot)$. Proposition 2 in Blundell et al. (2007) with data on wages, non-labor income, female labor supply and male labor participation allows the recovery of preferences and sharing rule up to an additive constant when $h_f > 0$. To estimate the bargaining effect of the paternity law on Costa Rican households, I split the sample according to those affected by the law and those that were not. Applying the identification for each sub-sample allows me to recover the sharing function parameters and compare them.

For the stochastic specification, I assume that the errors terms $(u_{1,i,t}, u_{0,i,t}, u^m_{p,i,t}, u^m_{w,i,t}, u^f_{w,i,t}, u_{y,i,t})$ are jointly conditionally normal with constant variance. Following Blundell et al. (2007), I include additive observed heterogeneity in the labor supply functions and the sharing rule. Additive heterogeneity ensures that the identification results of the sharing rule remain valid.

However, the heterogeneity might come from the labor supply or the bargaining function. For this reason, is that the constants in the structural equations are not identified.

I allow for general time effects in preferences and the sharing rule by including time dummies in the model.

2.D.5 Imputation and Likelihood

I impute wages for non-working spouses using a two-step Heckman selection estimation. I do this by first estimating a participation equation for both males and female ($j = m, f$):

$$\begin{aligned} p_{i,t}^{j} = \; & \beta_{0,t}^{j} + \beta_{1,t}^{j}\text{educ}_{i,t}^{f} + \beta_{2,t}^{j}\text{age}_{i,t}^{f} + \beta_{3,t}^{j}(\text{age}_{i,t}^{f})^{2} + \\ & \beta_{4,t}^{j}\text{educ}_{i,t}^{m} + \beta_{5,t}^{j}\text{age}_{i,t}^{m} + \beta_{6,t}^{j}(\text{age}_{i,t}^{m})^{2} + \beta_{y}^{j}y_{i,t} + \beta^{j} \cdot Y_{i,t}' + v_{i,t}^{j} \end{aligned} \tag{A14}$$

where Y are geographical and household variables. In the second step, I estimate the spouses' wage equations including the inverse Mills ratio. I impute wages using the predicted values. I impute non-labor income using the predicted values of equation A2.

After imputing wages and non-labor income, I estimate the structural model in two stages. The first stage is estimating the participation frontier for the husband, equation (2.12), with a probit. The second stage is estimating the wife's labor supply using a truncated regression. The likelihood function depends on the husband's labor participation. Define $f(\cdot)$ as the conditional normal density function and $\mathbf{1}(\cdot)$ as the indicator function. The likelihood when the husband works and there are n_W such observations are:

$$\begin{aligned} \log L^{W} = \; & \sum_{i=1}^{n_W} \{\mathbf{1}(h_{i,t}^{f} < 0) \log \Pr(p_{i,t}^{m} > 0, h_{i,t}^{f} < 0) + \\ & \mathbf{1}(h_{i,t}^{f} > 0)[\log \Pr(p_{i,t}^{m} > 0) + \log f(h_{i,t}^{f} | p_{i,t}^{m} > 0)]\} \end{aligned} \tag{A15}$$

The likelihood when the husband does not work and there is (n_N) such observations are:

$$\begin{aligned} \log L^{N} = \; & \sum_{i=1}^{n_N} \{\mathbf{1}(h_{i,t}^{f} < 0) \log \Pr(p_{i,t}^{m} < 0, h_{i,t}^{f} < 0) + \\ & \mathbf{1}(h_{i,t}^{f} > 0)[\log \Pr(p_{i,t}^{m} < 0) + \log f(h_{i,t}^{f} | p_{i,t}^{m} < 0)]\} \end{aligned} \tag{A16}$$

2.E Unitary model

The other common model of household behavior besides the collective model is the unitary model, which assumes that households behave as one single decision-maker.

For the unitary model, the household behaves as a single unit according to a twice continuously differentiable, strictly monotonic, and strongly concave utility function $U^H(l^f, l^m, C)$. It maximizes the following program:

$$\max_{h^f, h^m, C} \; U^H(l^f, l^m, C) \tag{A17}$$

$$s.t. \begin{cases} C = w^f h^f + w^m \overline{h}^m \mathbf{1}\{h^m = 0\} + y & \text{(A17a)} \\ 0 < h^f \leq 1, \quad h^m \in \{0, 1\} & \text{(A17b)} \\ l^f + h^f = 1 & \text{(A17c)} \\ l^m + h^m = 1 & \text{(A17d)} \end{cases}$$

In the unitary case, the household pools income, such that an increase in non-labor income or in the man's wage has the same effect on the woman's labor supply (only if the man works). The extra income has a standard income effect on the man's labor participation.

The unitary model uses the same parametric specification, providing two restrictions that be tested.

$$\begin{aligned} A_t &= A_y \\ a_t &= 0 \end{aligned} \quad \text{(U1)} \qquad \begin{aligned} (1+\gamma_y)(a_y - A_y) &= 0 \\ A_y \gamma_p &= (1+\gamma_y) a_p^* - A_p^* \end{aligned} \quad \text{(U2)}$$

The null hypothesis for these restrictions is that the household acts as a single unit decision-maker. Under the null, restrictions U1 refer to the household income distribution and how it affects the woman's labor supply. There is the same effect of an increase in the man's wage and non-labor income on the woman's labor supply. If the man does not work, there is no effect on his potential wage. Restrictions U2 refer to income and substitution effects caused by shifts in the woman's labor supply on the man's decision to work. Increases in non-labor income result in higher reservation wages, allowing the man to not work. The same holds for a raise in the woman's wage.

2.F Results paternity law on structural estimation

The following tables present the complete results of the collective household model for the treated and control samples.

Table A4: Unrestricted estimation results - control sample

	Woman's weekly labor hours				**Man's labor participation**	
	Man does not work		Man works			
	Coef	**SE**	**Coef**	**SE**	**Coef**	**SE**
Hourly wage man	-3.667	(2.430)	-1.405	(2.430)	1.037	(0.181)
Hourly wage woman	2.114	(4.472)	4.965	(2.184)	-0.060	(0.157)
Non-labor income	-4.688	(9.764)	0.113	(1.180)	-0.343	(0.332)
Intercept	35.762	(8.606)	39.684	(5.061)	-0.731	(0.262)
CV rural	1.979	(2.348)	-3.440	(2.158)	0.001	(0.096)
Non cv urban	2.542	(2.406)	-2.854	(2.066)	0.195	(0.100)
CV urban	4.297	(3.025)	-1.787	(1.998)	0.084	(0.100)
Children number	-1.589	(1.017)	-2.040	(0.774)	-0.047	(0.032)
Man has hs or college diploma	-0.066	(2.917)	-2.310	(1.853)	-0.151	(0.101)
Man's age	-0.266	(0.165)	-0.027	(0.117)	-0.010	(0.006)
Woman has hs or college diploma	2.609	(2.387)	3.025	(1.784)	0.079	(0.091)
Woman's age	0.456	(0.272)	0.145	(0.191)	0.013	(0.009)
Year effects	Yes		Yes		Yes	
N	1,376		2,622		3,998	

Standard errors are under parenthesis and have been computed using the bootstrap with 1000 repetitions and allowing for the fact that man's wage, woman's wage, and non-labor income are predicted.

Table A5: Collective estimation results - control sample

	Woman's weekly labor hours				**Man's labor participation**	
	Man does not work		Man works			
	Coef	**SE**	**Coef**	**SE**	**Coef**	**SE**
Hourly wage man	11.484	(0.337)	-2.641	(3.364)†	0.833	(0.121)
Hourly wage woman	10.387	(0.415)	9.341	(2.172)†	0.062	(0.133)
Non-labor income	-6.290	(0.031)	-6.284	(0.030)	0.000	(0.003)
Intercept	6.254	(0.135)	15.127	(0.450)	-0.633	(0.238)
CV rural	2.027	(0.053)	-2.577	(0.039)	0.006	(0.097)
Non cv urban	2.092	(0.030)	-1.045	(0.217)	0.175	(0.098)
CV urban	2.007	(0.081)	-1.189	(0.154)	0.091	(0.094)
Children number	-0.062	(0.697)	-2.497	(0.682)	-0.048	(0.032)
Man has hs or college diploma	-0.730	(0.329)	-1.196	(0.444)	-0.079	(0.087)
Man's age	-0.279	(0.157)	0.116	(0.115)	-0.008	(0.006)
Woman has hs or college diploma	0.394	(0.270)	2.708	(0.341)	0.054	(0.087)
Woman's age	0.661	(0.181)	0.688	(0.133)	0.011	(0.009)
Year effects	Yes		Yes		Yes	
N	1,376		2,622		3,998	

Standard errors are under parenthesis and have been computed using the bootstrap with 1000 repetitions and allowing for the fact that man's wage, woman's wage, and non-labor income are predicted.
†*: standard errors computed using the Delta Method.*

Table A6: Unrestricted estimation results - treated sample

	Woman's weekly labor hours				**Man's labor participation**	
	Man does not work		Man works			
	Coef	**SE**	**Coef**	**SE**	**Coef**	**SE**
Hourly wage man	6.557	(4.106)	-0.047	(4.106)	0.632	(0.232)
Hourly wage woman	14.584	(5.490)	8.937	(3.463)	-0.237	(0.191)
Non-labor income	-12.408	(15.343)	-22.812	(0.747)	-0.022	(0.433)
Intercept	31.105	(12.096)	45.158	(6.366)	-0.906	(0.475)
CV rural	-0.513	(3.332)	-6.814	(2.451)	0.135	(0.127)
Non cv urban	-7.654	(3.875)	-5.646	(2.404)	0.119	(0.124)
CV urban	-3.657	(4.754)	-5.424	(2.378)	0.218	(0.133)
Children number	1.522	(1.335)	0.420	(0.949)	-0.121	(0.043)
Husband has hs or college diploma	-2.028	(3.377)	-0.945	(2.495)	0.297	(0.126)
Husband's age	0.035	(0.207)	-0.439	(0.157)	-0.032	(0.007)
Wife has hs or college diploma	-4.506	(3.709)	1.057	(2.380)	0.177	(0.120)
Wife's age	-0.327	(0.422)	-0.096	(0.237)	0.037	(0.012)
Year effects	Yes		Yes		Yes	
N	1,506		2,868		4,374	

Standard errors are under parenthesis and have been computed using the bootstrap with 1000 repetitions and allowing for the fact that man's wage, woman's wage, and non-labor income are predicted.

Table A7: Collective estimation results - treated sample

	Woman's weekly labor hours				**Man's labor participation**	
	Man does not work		Man works			
	Coef	**SE**	**Coef**	**SE**	**Coef**	**SE**
Hourly wage man	9.633	(1.204)	-0.398	(2.941)†	0.640	(0.178)
Hourly wage woman	7.506	(1.681)	10.365	(2.194)†	-0.182	(0.138)
Non-labor income	-21.294	(1.299)	-21.472	(2.303)	0.011	(0.223)
Intercept	23.304	(3.837)	47.059	(4.592)	-0.920	(0.454)
CV rural	0.606	(1.337)	-7.049	(1.347)	0.133	(0.126)
Non cv urban	-6.584	(1.601)	-5.937	(1.499)	0.111	(0.121)
CV urban	-3.057	(1.041)	-5.559	(1.268)	0.210	(0.128)
Children number	1.403	(1.243)	0.437	(0.912)	-0.119	(0.041)
Husband has hs or college diploma	-2.815	(1.152)	-0.918	(1.910)	0.292	(0.113)
Husband's age	-0.014	(0.197)	-0.438	(0.147)	-0.032	(0.006)
Wife has hs or college diploma	-1.133	(1.231)	0.245	(1.735)	0.154	(0.106)
Wife's age	-0.059	(0.315)	-0.147	(0.216)	0.036	(0.011)
Year effects	Yes		Yes		Yes	
N	1,506		2,868		4,374	

Standard errors are under parenthesis and have been computed using the bootstrap with 1000 repetitions and allowing for the fact that man's wage, woman's wage, and non-labor income are predicted.
†*: standard errors computed using the Delta Method.*

2.F.1 Unitary model estimation

I can test the unitary model restrictions presented above. The null hypothesis for the unitary model is that the household acts as a single unit decision-maker. Testing the unitary restrictions with the treatment sample, the likelihood ratio statistic for the unitary model restrictions is 2.80 with a p-value of 0.59; hence, I do not reject that treated households behave as the unitary model predicts For the control sample, the likelihood ratio statistic for the unitary model is 0.23 with a p-value of 0.99; hence, I do not reject that control households behave as the unitary model predicts.

Table A8: Unitary estimation results - control sample

	Woman's weekly labor hours				**Man's labor participation**	
	Man does not work		Man works			
	Coef	**SE**	**Coef**	**SE**	**Coef**	**SE**
Hourly wage man	0.000		-1.258	(1.520)	1.036	(0.182)
Hourly wage woman	3.044	(1.488)†	4.412	(1.777)	-0.061	(0.154)
Non-labor income	-1.258	(1.520)	-1.258	(1.520)	-0.363	(0.335)
Intercept	32.245	(5.985)	39.878	(4.643)	-0.730	(0.260)
CV rural	1.559	(2.008)	-3.396	(2.020)	0.002	(0.089)
Non cv urban	2.308	(1.939)	-2.803	(1.866)	0.196	(0.096)
CV urban	3.841	(2.620)	-1.710	(1.833)	0.084	(0.097)
Children number	-1.537	(1.033)	-2.031	(0.734)	-0.047	(0.032)
Man has hs or college diploma	-1.464	(1.755)	-2.315	(1.560)	-0.151	(0.100)
Man's age	-0.296	(0.156)	-0.029	(0.115)	-0.010	(0.006)
Woman has hs or college diploma	2.040	(1.583)	3.224	(1.582)	0.079	(0.091)
Woman's age	0.457	(0.266)	0.149	(0.188)	0.013	(0.009)
Year effects	Yes		Yes		Yes	
N	1,376		2,622		3,998	

Standard errors are under parenthesis and have been computed using the bootstrap with 1000 repetitions and allowing for the fact that man's wage, woman's wage, and non-labor income are predicted.
†: standard errors computed using the Delta Method.

Table A9: Unitary estimation results - treated sample

	Woman's weekly labor hours				**Man's labor participation**	
	Man does not work		Man works			
	Coef	**SE**	**Coef**	**SE**	**Coef**	**SE**
Hourly wage man	0.000		-4.380	(3.732)	0.608	(0.220)
Hourly wage woman	6.530	(5.480)†	11.184	(3.746)	-0.220	(0.188)
Non-labor income	-4.380	(3.732)	-4.380	(3.732)	0.212	(0.212)
Intercept	36.638	(16.845)	49.337	(5.636)	-0.872	(0.466)
CV rural	0.052	(6.418)	-6.774	(2.391)	0.135	(0.124)
Non cv urban	-7.020	(3.409)	-5.703	(2.368)	0.118	(0.128)
CV urban	-1.819	(5.471)	-5.361	(2.355)	0.218	(0.135)
Children number	1.256	(1.289)	0.361	(0.932)	-0.122	(0.042)
Husband has hs or college diploma	0.269	(1.914)	0.749	(2.691)	0.304	(0.125)
Husband's age	0.055	(0.255)	-0.374	(0.154)	-0.032	(0.007)
Wife has hs or college diploma	-4.623	(2.847)	-0.036	(4.198)	0.167	(0.119)
Wife's age	-0.312	(0.598)	-0.214	(0.235)	0.036	(0.012)
Year effects	Yes		Yes		Yes	
N	1,506		2,868		4,374	

Standard errors are under parenthesis and have been computed using the bootstrap with 1000 repetitions and allowing for the fact that man's wage, woman's wage, and non-labor income are predicted.

†: standard errors computed using the Delta Method.

Chapter 3

Attrition and time-varying unobserved heterogeneity in an education-labor dynamic discrete choice model

with François Poinas.

Abstract

Understanding individuals' educational choices and their impact on outcomes like wages and job mobility is essential for policymakers. Using data from the French *Génération 98* survey, we propose a dynamic human capital accumulation model with two methodological improvements to examine the robustness of empirical data. We incorporate time-varying unobserved heterogeneity and endogenous attrition. Preliminary results show that attrition bias has no significant effect on estimated parameters. But it does affect sample composition over longitudinal panels and simulations by over-representing more-educated individuals with higher wages at older ages. The findings also show time-varying components in unobserved heterogeneity over time, in particular by accumulating schooling. It asses the importance of educational system tracking in skill accumulation and decision-making processes. We shed light on the diverse effects of accumulated schooling in skill acquisition and correcting for attrition in simulations that can be used for policy recommendations.

3.1 Introduction

Understanding individuals' schooling decisions and their impact on outcomes such as earnings and job mobility is crucial for policymakers who wish to design the right educational policies. To analyze schooling and career choices, economists typically employ a human capital framework, treating individuals as economic agents who can accumulate skills (human capital) through education or work. Following the seminal paper of Keane and Wolpin (1997), empirical economists have estimated dynamic structural models of human capital accumulation. This approach consists of modeling explicitly individual decisions about attending school and working (considering, possibly, working in different occupations) during the life cycle. At each age, agents take the decision that maximizes their life-time satisfaction, considering that their present decision affects subsequent decisions and outcomes. The parameters of the behavioral model are estimated using observational data on individual schooling attainments, labor market transitions, and wages in a revealed preference approach. As these models rely on several assumptions, it is essential to assess the robustness of the empirical results to alternative assumptions.

An essential aspect of these models is the inclusion of time-unvarying unobserved heterogeneity, recognizing that individuals possess comparative advantages in skill accumulation, analogous to a Roy-type model. These unobserved traits, which encompass cognitive skills (e.g., intelligence) and non-cognitive skills (e.g., motivation, personality traits), have a direct impact on educational attainment and labor market success. A growing empirical literature shows that these skills evolve with age and can be altered by education (see Cunha et al., 2010; Todd and Zhang, 2020). Another aspect that has been overlooked in estimating dynamic models of human capital accumulation is related to missing observations in the datasets due to attrition, even though many of the longitudinal datasets used in the literature have a significant level of attrition. In the "National Longitudinal Survey of Youth 1979" (National Longitudinal Surveys, 2023), the sample loses 33% of observations between 1979 and 2014 (Rothstein et al., 2018), and in the "Household, Income, and Labour Dynamics in Australia Survey" (HILDA), the sample loses close to 40% of respondents between the first and sixteen waves. If observations leaving the sample were drawn randomly, this aspect would be innocuous in estimating the model's parameters. However, if the reasons for leaving the sample were endogenous, i.e., linked to individual characteristics that affect schooling choices and educational outcomes or directly linked to the choices and outcomes, the omission of these observations would lead to a selection bias on the estimates.

In this paper, we estimate a dynamic human capital accumulation model that incorporates two methodological improvements. First, we incorporate time-varying unobserved heterogeneity when, unlike Todd and Zhang (2020), the econometrician does not have access to time-varying variables measuring skills and needs to rely on standard data sets containing observations of school choices, employment, wages, and some observed characteristics only. Second, we correct for a potential attrition bias by incorporating explicitly a selection mechanism into the model. We estimate our model with *Génération 98* data, a French survey that contains rich information about education trajectories and labor market outcomes for a large sample of individuals who finished education in 1998 and were followed for 10 years.

Incorporating unobserved heterogeneity allows controlling for the ability bias, enabling the identification of causal effects, such as the returns to schooling. Furthermore, it allows to simulate the effect of counterfactual policy changes, such as increasing the minimum school-leaving age or altering higher education tuition fees. By assuming that an individual's unobserved ability remains stable over time, the standard human capital accumulation model described above does not explicitly acknowledge this possibility. Consequently, it rules out the possibility that higher accumulated schooling can increase the aptitude of the individual to accumulate skills at school. Ignoring this channel can be criticized for its lack of realism, especially in educational systems in which students are sorted, at different ages into different tracks, some of them being designed explicitly to make students more effective at acquiring learning skills at school.

Allowing for this type of behavior seems to be particularly important in the French context, where, like in many continental European countries, students are tracked into different programs at the secondary education level, that differ in terms of content and objective (Belzil and Poinas, 2018). Attending a general track in secondary education (compared to a vocational track) may, at the same time, increase educational attainment, and give skills that make the individual more able to accumulate more schooling later (in higher education). In the model we propose, accumulated years of schooling affect both directly the labor market productivity and make the individual more likely to increase the individual's comparative advantage of accumulating skills at school.

The data we use, *Génération 98*, is characterized by the presence of attrition: by the fourth wave of data collection in 2008, 70.45% of individuals surveyed in the first wave have left the sample. Moreover, descriptive statistics show that attrition is related to individual characteristics and choices. Therefore, controlling for endogenous attrition seems to be of particular importance. In our model, we incorporate attrition as an additional choice that the agent faces at the time of an interview. This attrition choice becomes endogenous by depending on previous choices and unobserved heterogeneity. This way of making attrition endogenous can be adapted to similar datasets where attrition is prevalent.

To assess the role played by both the introduction of time-varying unobserved heterogeneity and endogenous attrition, we estimate different versions of a human capital accumulation model. In the model, individuals choose, from age 16 to 35, between three options: attending school, working, and staying home. When individuals enter the labor market, they get utility from labor market wages. Individuals' utilities and wages are affected by unobserved individual-specific characteristics, accumulating schooling, and labor market experience. The benchmark specification is a standard model with time-unvarying unobserved heterogeneity that does not control for attrition. Other versions incorporate the aspects related to time-varying unobserved heterogeneity and endogenous attrition described above. By comparing parameter estimates and counterfactual simulation results obtained with the different specifications, we can quantify the potential biases resulting from the omission of our proposed methodological contributions.

To illustrate the role of each of these two components, we consider the anticipated effect of the counterfactual policy of raising tuition fees. In a model with time-varying unobserved heterogeneity, on top of the direct effect the policy has on the cost of acquiring schooling, it might also have an indirect effect by reducing the ability to accumulate skills in school, i.e., decreasing the probability to move from a low schooling ability type to a high type. By decomposing the total effect into a direct and an indirect effect, it permits shedding light on the heterogeneous effects of the policy and assessing at which stage of the schooling accumulation process the policy plays. Concerning endogenous attrition, if lower productive ability types are more likely to leave the sample earlier, not controlling for it is expected to lead to an upward bias in the way wages evolve with experience (as individuals with a high productive type would be over-represented in the sample with large years of experience). This would result in an overestimation of schooling attainments. Correcting this bias would alter the simulated effect of the tuition fee policy.

Preliminary results show that attrition bias has no significant effect on the model's estimated parameters. One possible explanation is that accounting for unobserved heterogeneity already corrects for attrition. However, we find a negative and significant correlation between accumulated years of experience and schooling and the probability of an individual leaving the sample. This means that simulations that do not account for attrition may have different types of older individuals than models that do, with more educated and work-experienced people remaining in the data. This diverse sample has an impact on longitudinal panels and simulations, that are then utilized to create policy recommendations. In the case of time-varying unobserved heterogeneity, our findings indicate a time persistence difference in the transition probability of unobserved heterogeneity. Schooling years improve the probability of being of a type with more rewarding benefits for later-life decisions. This is a new indirect effect of human capital accumulation on individuals. We also discover that the probability of individuals changing types diminishes with age, approaching zero at the age of 30.

3.1.1 Related literature

Our paper contributes to two different literatures. First, it contributes to the literature on time-varying unobserved heterogeneity by providing an empirical estimation of a dynamic discrete choice model with time-varying unobserved heterogeneity without the use of exogenous time-varying variables. In a reduced-form setting, Ding and Lehrer (2014) estimate an education production function by allowing time-varying unobserved heterogeneity. They motivate their study by arguing how, since human capital accumulation is a dynamic process, it is important to understand how the role of heterogeneous ability evolves over the life cycle and hence, the necessity of allowing for time-varying in the unobserved heterogeneity term of the education's production function. Their result points out how different conclusions can be reached depending on the assumption impose on unobserved ability heterogeneity. In the structural literature, Arcidiacono and Miller (2011) introduced an extending form of the conditional choice probability estimator that allows for an unobserved time-varying state in the model. However, they do not perform an empirical estimation. Hu and Shum (2012) presents the identification of the law of motion governing the time-varying unobserved heterogeneity in a stationary and non-stationary structural setting. To our knowledge, there is only one empirical study allowing for time-varying unobserved heterogeneity: Todd and Zhang (2020). By extending the proof in Hu and Shum (2012) in an age-dependent setting they introduce time-varying unobserved heterogeneity in á la Keane and Wolpin (1997) model. Their main result is that allowing personality traits to evolve with age and with schooling proves to be important to capture the heterogeneity in how people respond to educational policies. We contribute to the literature by providing another empirical study without the use of external time-varying variables but only state variables.

The second literature we contribute is on attrition. Attrition is one of the main problems in panel datasets. Moffit et al. (1999) follows Heckman's selection mechanism to model how attrition is an endogenous process, especially from lagged dependent variables. They call it "selection on observables". In their application to the Michigan Panel Study of Income Dynamics, they find attrition bias in the earnings of male heads of households and that attrition is selective on the stability of earning profiles. They also show differences in socioeconomic variables for people who left the sample. For example, the male head of households who left the sample was more likely to benefit from subsidies, they were less likely to marry, they have lower levels of education, they worked for fewer hours, and they were older. The main mechanisms for the selection bias to happen are by correlation in the output shocks with the selection equation shock or through unobserved heterogeneity affecting outputs and selection. Kyriazidou (2001) provides identification of a dynamic model with a dynamic selection process. He considers unobserved heterogeneity as the way the attrition bias is present and introduces a difference estimator that eliminates the heterogeneity. Her method has been widely used in the literature since then. Gayle and Viauroux (2007) present a dynamic panel-sample-selection model where both the outcome and selection equations are affected by lag values of the outcome variable as well as unobserved effects. Lastly, Alfo and Maruotti (2009) presents a model with sample selection affected by unobserved heterogeneity. To correct the attrition bias, they use a finite mixture model to estimate the unobserved heterogeneity in both the outcome and selection equation. Their main assumption is that the outcome and selection variables are independent, conditional on the unobserved effect. Spagnoli et al. (2018) follow this approach but introduce a bi-dimensional finite mixture to help create a direct relationship between ignorable and non-ignorable selection and a way to test for it. We contribute to this literature by introducing a simple sample correction for attrition endogenous to the model. We allowed attrition to be corrected by exogenous variables and unobserved heterogeneity.

The paper is structured as follows: The second section describes the data we use. Section three then introduces the model and discusses the model identification. Section four shows the model's solution and estimation. Section five presents the model's preliminary results. Finally, section six concludes.

3.2 Data

We use *Génération 98*, a French suvery[1]. It follows individuals that finished their education in 1998, at any level of education, for 10 years. Information was collected in 4 different waves. The first wave of interviews was performed in 2001 on 55,345 individuals.

Our sample consists of 27,303 men[2] aged between 16 and 30 years old who finished their education in 1998. We do further sample selection choices that are explained in Appendix 3.A.

The data collects rich information on demographics and family background characteristics. Table 3.1 presents descriptive statistics of the variables used in the analysis for the sample observed in 1998. The sample is composed of individuals aged 21 years old on average. 84% of them live in an urban location and 79% have French parents. 44% (resp. 58%) of individuals

[1] More information about *Génération* 98 is available at `https://data.progedo.fr/studies/doi/10.13144/lil-0167`.

[2] Females are not included to avoid modeling fertility decisions.

have fathers (resp. mothers) working in high-skilled jobs (executive or white collar).

Table 3.1: Descriptive Statistics in 1998

Variable	Mean	Std. Dev.	Min	Max
Age	21.32	2.92	16	30
Urban location	0.84		0	1
Parents nationality:				
Both parents French	0.79		0	1
One parent French	0.09		0	1
None parent French	0.13		0	1
Late in middle school	0.27		0	1
Father's occupation:				
No occupation or farmer	0.24		0	1
Technical or blue collar	0.32		0	1
Executive or white collar	0.44		0	1
Mother's occupation:				
No occupation or farmer	0.27		0	1
Technical or blue collar	0.15		0	1
Executive or white collar	0.58		0	1
Number of observations	27,303			

3.2.1 Sample Attrition

Among our initial sample of 27,303 individuals who appear in the first wave of interviews (2001), 16,297 individuals were randomly selected to be interviewed in the second wave (2003). 10,469 of these individuals answered the interview. For the following waves, all individuals were eligible for a re-interview. Among them, 7,282 individuals answered the survey in the third wave (in 2005), and 4,816 individuals answered the survey in the fourth wave (in 2008). Table 3.2 shows the attrition rate of the sample randomly selected in the second wave across the four waves. Attrition rates from one wave to the other are of high importance, accounting for 70% from the first to the fourth wave.

Table 3.2: Sample Attrition

Period	# of observations	% Attrition (previous period)	% Attrition (total)
1998-2001			
Full sample	27,303		
Random sample	16,297		
2002-2003	10,469	35.76%	35.76%
2004-2005	7,282	30.44%	55.32%
2006-2008	4,816	33.86%	70.45%

Note: The random sample corresponds to the individuals who were randomly selected to be re-interviewed in the second wave (2003). Attrition rates are calculated on the random sample.

We model attrition affecting the random sample as endogenous, as detailed in the following section. We use variables that are included in the attrition equations but exclude from the utility associated with schooling and working decisions. These exclusion restrictions are likely to influence a respondent's decision to participate in the survey (assuming they are randomly selected in the original sample in wave 2), but they have no effect on school or employment decisions. We consider variables related to the family situation at the time of the attrition event (a dummy of being in a couple and dummies associated with the number of children in the family) as well as a variable measuring the distance between the location in middle school (at age 11) and the location at the time of the attrition event. These characteristics are likely to affect attrition by changing the cost of responding to the survey or the survey institution's capacity to reach the respondent.

To illustrate the correlation of attrition with endogenous variables and the exclusion restriction variables presented in the previous paragraph, we regress the attrition dummy event on years of schooling, work experience, family, and distance variables as well as control variables. Results are presented in Table 3.3. More educated individuals are less likely to leave the sample as well as individuals with more work experience. The family situation is also correlated to attrition (having children decreases the probability to leave the sample), after controlling for education and experience.

Table 3.3: Attrition Correlations

	(1) Wave 1-2	(2) Wave 2-3	(3) Wave 3-4
Years of schooling	-0.0111***	-0.0108***	-0.0166***
	(0.00155)	(0.00194)	(0.00273)
Work experience	-0.0225***	-0.00930**	-0.0121***
	(0.00343)	(0.00286)	(0.00282)
In couple	0.0126	0.0274**	0.0172
	(0.00995)	(0.0100)	(0.0122)
One child	-0.0970***	0.0148	-0.170***
	(0.0161)	(0.0148)	(0.0177)
Two or more children	-0.0871*	0.00516	-0.0633**
	(0.0365)	(0.0244)	(0.0211)
Urban location	0.0109	0.0274*	0.0512***
	(0.00989)	(0.0114)	(0.0141)
Distance to middle school	0.00566*	0.00428	-0.00288
	(0.00280)	(0.00312)	(0.00391)
Constant	0.590***	0.467***	0.679***
	(0.0277)	(0.0340)	(0.0464)
Controls	Yes	Yes	Yes
Observations	16,297	10,469	7,282
R^2	0.014	0.009	0.035

Note: standard errors in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Control includes the parents' profession, nationality, and being late to middle school.

3.2.2 Choices over Ages

We are interested in modeling individual decisions between staying at school, working, and staying home. The "home" option regroups the different situations that are neither school nor work, like deciding to stay out of the labor market or being unemployed. From the monthly observations available in the data, we construct yearly decisions at each age, from age 16 to 35. Appendix 3.A provides details about the construction of these decisions. Figure 3.1 represents the distribution of choices over ages. At age 16, almost all individuals are still in school.[3]

[3] The minimum age to leave school is set to 16 years old in France.

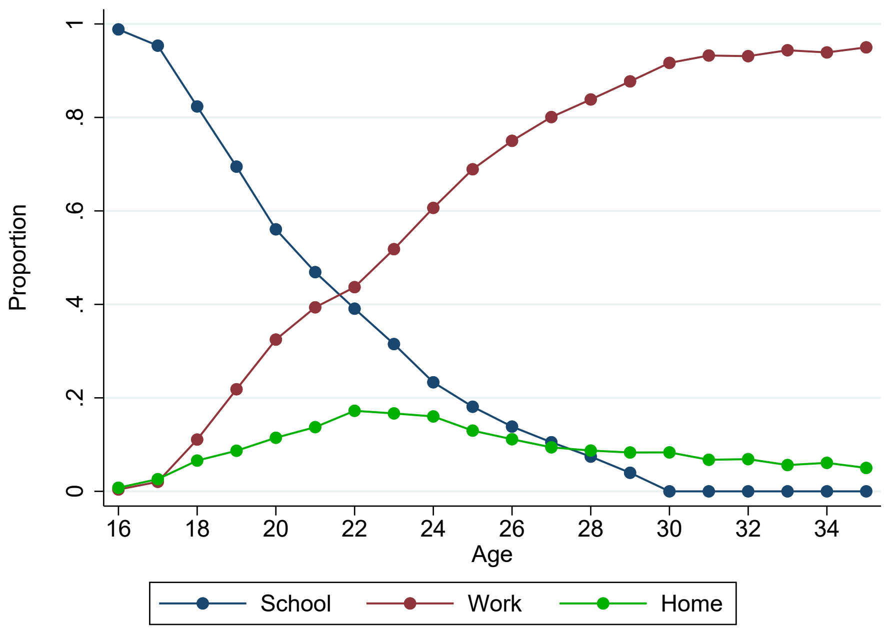


Figure 3.1: Choices by Age

Starting at age 18, the fraction of individuals leaving school increases. Most individuals enter the labor market, but a significant proportion stays home. At age 22, which corresponds to the age of reaching an undergraduate level without any grade interruption, the school and work curves cross each other, and it corresponds to the age at which the proportion of individuals staying home is the largest, at almost 20%. Given we restrict our sample to individuals younger than 30 years old in 1998, the fraction of individuals who are still in school at that age is 0 by construction.

Table 3.4 presents the one-year transition rates. The *Génération 98* survey collects information on individuals who finish their education in 1998. By construction, we do not observe individuals re-entering school. Given the young ages we are considering in the model, the high state dependency on the school option (81%) is not surprising. State dependency on working is also extremely large (95%), whereas the proportion of individuals who make a transition from home to work is around half.

Table 3.4: Choice Transition Matrix

Choice	Choice at age t			
at age $t-1$	School	Work	Home	Total
School	81.43%	10.97%	7.61%	100.00%
Work	0.00%	95.13%	4.87%	100.00%
Home	0.00%	49.00%	51.00%	100.00%

3.3 Model

We develop a decision model for men to choose between three mutually exclusive alternatives $j \in J$: school ($j = 1$), work ($j = 2$), and staying at home ($j = 3$), which also includes unemployment, between ages 16 and 29. From age 30 to 64, they choose only between work and staying at home. At each age, individuals maximize their remaining discounted lifetime utility. The terminal age is 65, but given we only observe individuals in their early years, we assume that individuals make choices until age 35 and then stay in their age 35 choices up to the age of 65.

At the age of 16, individuals have an initial endowment determined by family background characteristics (parents' nationality and profession), and if the individual was older than the normal age to enter middle school.[4] To allow for unobserved heterogeneity, we assume individuals can be one of two types $k = \{1, 2\}$. The pecuniary and nonpecuniary rewards that an individual receives depend on their type and the choices that they select.

We allow individuals to change type when their age increases, following Todd and Zhang (2020). The probability to switch to a new type varies with age and endogenous choices.[5] At age 16, the initial type $k(16)$ is determined by the initial endowments $s_0(16)$. Given the initial type and initial endowments, the individual chooses the alternative $d_j(16)$ that gives the highest continuation value. At the next periods, the state variables update according to the choice made at the previous period $d_j(t-1)$ and the new type $k(t)$, which itself depends on the observed state variables and the previous type $k(t-1)$.

3.3.1 State variables and Laws of Motion

The time-varying part of the state space consists of $g(t)$, the individual's accumulated years of schooling at age t, $x(t)$, the accumulated years of work experience, and the individual's type $k(t)$. $s(t)$ denotes the observed state variables, i.e., $s(t) = \{g(t), x(t)\}$.

Years of schooling and work experience evolve in a deterministic way. The updating proceeds as follows:

$$
\begin{aligned}
g(t+1) &= g(t) + 1, \text{ if } d_1(t) = 1 \\
x(t+1) &= x(t) + 1, \text{ if } d_2(t) = 1\,.
\end{aligned}
$$

[4] In France, the average age to enter middle school, or *Sixième*, is 11 years old.

[5] In Todd and Zhang (2020), endogenous choices (accumulated schooling) affect the probability to switch to a new type through the effect it has on the individual's personality traits. In our setup, we do not introduce personality traits (as we do not observe them), so choices directly affect the probability to switch to a new type.

Following Todd and Zhang (2020), the evolution of the individual's type follows a Markovian process. After the initial period, at each age t, the type $k(t)$ has a probability $p(t)$ of changing or $1 - p(t)$ of remaining the same. Conditional on changing, $q_k(t)$ represents the probability of becoming type $k = \{1, 2\}$ at age t.

Let $M(t)$ be the type transition matrix between period t and $t + 1$ for two types.[6] It takes the following form:

$$M(t) = \big(1 - p(t)\big) \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} + p(t) \begin{bmatrix} q_{k=1}(t) & q_{k=1}(t) \\ q_{k=2}(t) & q_{k=2}(t) \end{bmatrix} ,$$

where $p(t)$ is the probability of changing type at period t. We define this probability as a function of the age of the individual. It is defined as:

$$p(t) = \frac{1}{1 + \exp(\varphi_1 + \varphi_2(t - 16))} . \tag{3.1}$$

We assume $q_k(t)$ follows a multinomial logit form and it depends on two sets of variables: time-unvarying and time-varying. Time invarying characteristics (X_{16}) include family background characteristics (parents' occupation and nationality) and a dummy for being late at middle school entry. We use accumulated years of schooling ($g(t)$) as the time-varying component. $q_k(t)$ takes the following form:

$$q_k(t) = \frac{\exp(v_k(t))}{\sum_{k=1}^{K=2} \big(\exp(v_k(t))\big)}$$

$$v_k(t) = \varphi_{4,k} + \varphi_{5,k} X_{16} + \varphi_{6,k} g(t) + \xi_k(t) ,$$

where φ's are parameters[7] and $\xi_k(t)$ follows a type-1 extreme value distribution.

3.3.2 Alternatives' Utilities

At each year of education, the individual receives a utility which is composed of a nonpecuniary component, which captures the consumption value, or any physical or mental costs of attending a year of education, and a pecuniary component, which represents, for instance, tuition or housing costs. The utility of attending education at time t is:

$$u_1(t) = \theta_s(k) + \beta_0 \text{urban}(t) + \beta_1 \mathbf{1}\{\text{age} \geq 18\} + \varepsilon_s(t) ,$$

[6] The Markov process for the evolution of types can be generalized to K different types. The selection of the number of types is described further below.

[7] For identification, we fix parameters for type two to 0.

where urban(t) is a dummy indicating if the individual is living in an urban area at age t, $\beta_1 \mathbf{1}\{\text{age} \geq 18\}$ captures the cost of attending school after the age of 18. $\theta_s(k)$ is type k's specific reward of attending school and lastly, $\varepsilon_s(t)$ is a preference shock.

The utility of working is composed of the wage $w(t)$, a non-pecuniary reward $r(t)$ (not separately identified from the type-specific intercept), and a preference shock $\varepsilon_w(t)$:

$$u_2(t) = w(t) + r(t) + \varepsilon_w(t)\,. \tag{3.2}$$

Following Keane and Wolpin (1997), the wage is specified as a human capital pricing equation. It is the product of the price per unit of human capital p and the amount of human capital $e(t)$. Hence, $w(t) = p\,, e(t)$. The amount of human capital is determined by type k's comparative advantage $\theta_w(k)$, if the individual is living in an urban area at time t, accumulated years of schooling $g(t)$, work experience $x(t)$, and a dummy for the first two years an individual gain work experience. The respective log-wage equation form is:

$$\log w(t) = \log(p) + \theta_w(k) + \gamma_0 \text{urban}(t) + \gamma_1 g(t) + \gamma_2 x(t) + \gamma_3 \mathbf{1}\{x(t) \leq 2\} + \eta(t)\,,$$

where $\eta(t)$ is a skill technology shock, assumed to be i.i.d. normal distributed with variance σ_w^2. For identification, we fix the nonpecuniary reward $r(t)$ to zero.

Lastly, the utility of staying at home is composed of the type-specific component $\theta_h(k)$, an age effect, and the preference shock $\varepsilon_h(t)$:

$$u_3(t) = \theta_h(k) + \delta_1 age + \varepsilon_h(t)\,.$$

3.3.3 Attrition

In the re-interview years (2003, 2005, and 2008), a fraction of individuals leaves the sample. To model attrition endogenously, we treat the fact that an individual stays in the sample as an individual decision to participate in the survey at the re-interview date. As already explained in Section 3.2.1, only a sub-sample was randomly drawn from the initial sample to be interviewed in the first re-interview year (2003). We treat the attrition results from this random draw as exogenous (i.e., we simply consider that subsequent information is missing for individuals not randomly selected) and model the decision to participate or not in the 2003 interview as endogenous only for the individuals who are in the randomly selected sample.

The decision to stay in the sample happens at certain individual-specific ages, which depends directly on the age in 1998. We assume that the decision depends on accumulated years of schooling and experience, a type-specific component $\theta_a(t)$ (to make it depend on individual specific attributes not observed by the econometrician), and on individual observed characteristics $X_a(t)$ (couple situation, number of kids, urban residency and distance between the location in middle-school and the location measured in the last survey). The decision to leave the sample takes then the following form:

$$\begin{aligned} a(t)^* &= \theta_a(t) + \psi_1 g(t) + \psi_2 x(t) + \boldsymbol{\psi} X_a^{'}(t) + \varepsilon_a(t) \\ a(t) &= \mathbf{1}\{a(t)^* > 0\}\,, \end{aligned}$$

where $\varepsilon_a(t)$ is an i.i.d. preference shock that follows a type-1 extreme value distribution.

3.3.4 Information structure and timing

The timing goes as follows: at the beginning of each period, each individual realizes his type for the period t. Then, the individual observes the preference shocks and decides. After the decision is taken, he observes the wage. If period t includes an attrition decision, the attrition shock happens right after he knows his type.

An individual in state s knows all state variable laws of motion, $Pr(s(t+1)|s(t), d_1(t), d_2(t), d_3(t))$. He uses the distribution of wage shocks $F_w(\eta(t))$, idiosyncratic preference shocks $F_j(\varepsilon_j(t))$, attrition shock $F_a(\varepsilon_a(t))$, and type transition shocks $F_k(\xi(t))$ to form an expectation over future states. For computational simplicity, $\eta(t)$ is normally distributed with mean 0 and variance σ_w^2, whereas $\varepsilon_j(t)$, $\varepsilon_a(t)$ and $\xi(t)$ are assumed to be type-1 extreme value distributed. Conditional on the unobserved types, the other shocks are assumed to be i.i.d. over time.

3.3.5 Identification

Time-unvarying multinomial types were introduced in a career decision model by Keane and Wolpin (1997) to control for unobserved heterogeneity. Hu and Shum (2012) provide the first identification's proof for time-varying unobserved heterogeneity in a dynamic model. The main assumption is limited feedback, which relies on two conditions: (1) the decision made at each period t is independent of the decision made the period before $t-1$, after conditioning on the state variables at period t, and (2) the type $k(t)$ is independent of last period's decision after conditioning on the other state variables at t and the state variables at $t-1$. Todd and Zhang (2020) generalize the identification's proof to an age-dependent setting where only three time periods are needed (Hu and Shum (2012) mention the necessity of at least four time periods). The complete proof is in Todd and Zhang (2020), Appendix A.

The identification of the structural parameters goes as follows. In the last period of the model, the choice between the different alternatives is analogous to that of a multinomial logit model, given the types. Because of the normalization of the nonpecuniary reward in work utility and wage data, it is possible to recover the wage parameters (the price per unit of human capital p_{hc} cannot be identified separately from the unobserved heterogeneity term in the estimation). With the wage parameters identified, it is possible to identify the utility parameters. Following Arcidiacono (2005) and Beffy et al. (2012), we use exclusion restrictions to control for selection into each decision. Joint with the assumed functional form of the model, it is possible to identify the unobserved characteristics.

The identification of the attrition parameters relies on the assumption that attrition is conditionally random as defined by Little and Rubin (2019). Following Alfo and Maruotti (2009), conditioning the attrition decision on a variable like unobserved heterogeneity, the attrition decision becomes random and independent from the other decisions. With attrition defined as conditionally random, it is possible to separate the joint probability of the outcome variables and attrition and identify the parameters of the attrition equation. The exclusion restriction variables we use are the distance between the interview location and middle school location, used as a proxy for the distance to hometown, and personal characteristics of the individual when the attrition decision was present as (marital status and the number of kids).

3.4 Estimation

3.4.1 Solving the Model

Each individual starts age t with an observed state vector $s(t)$ and type $k(t)$. Denote $d_j(t) = 1$ when alternative j is chosen at time t. The value function at age t of an individual of type $k(t)$ is the maximum over all possible sequences of future decisions given the current state space and the set of parameters Ω:

$$V(s(t), k(t), \Omega) = \max_{\{d_j(t)\}} \mathbb{E}\left[\sum_{\tau=t}^{T} \delta^{\tau-t} \sum_{j=1}^{J} u_j(t) d_j(t) \;\middle|\, s(t), k(t) \right].$$

The expectation is taken over future wages, preference shocks, and the unobserved type's transition process. The value function can be written in a Bellman equation form, where for each possible decision j:

$$\begin{aligned} V_j(s(t), k(t), \Omega) &= u_j(t) + \delta\, \mathbb{E}\big[V(s(t+1), k(t+1), \Omega) \mid s(t), k(t), d_j(t)\big] \quad \text{for } j = 1, 3 \\ V_j(s(t), k(t), \Omega) &= \widetilde{u}_j(t) + \delta\, \mathbb{E}\big[V(s(t+1), k(t+1), \Omega) \mid s(t), k(t), d_j(t)\big] \quad \text{for } j = 2\,, \end{aligned}$$

where $\widetilde{u}_j(t) = \int_{\eta(t)} u_j(t) f(w(\eta(t))) d\eta(t)$ is the utility after integrating over the wage shock distribution.

Let $\widetilde{V}_j(s(t), k(t), \Omega)$ the choice-specific value function excluding the contemporaneous decision-specific preference shock $\varepsilon_j(t)$:

$$V_j(s(t), k(t), \Omega) = \widetilde{V}_j(s(t), k(t), \Omega) + \varepsilon_j(t)\,.$$

As the preferences shocks follow an i.i.d. type-1 extreme value distribution, one has the closed form solution for the expectation (Rust, 1987). The expectation of the value function can then be written as:

$$\begin{aligned} &\mathbb{E}\big[V(s(t+1), k(t+1), \Omega) \mid s(t), k(t), d_j(t)\big] \\ = \;& \mathbb{E}_{k(t+1), \varepsilon_j(t+1)} \left[\max_{d_j(t+1)} \sum_{j=1}^{J} d_j(t+1) \Big\{ \widetilde{V}_j(s(t+1), k(t+1), \Omega) + \varepsilon_j(t+1) \Big\} \mid s(t), k(t), d_j(t) \right] \\ = \;& \mathbb{E}_{k(t+1)} \left[\underbrace{\lambda \log \sum_{j=1}^{3} \exp\left(\frac{\widetilde{V}_j(s(t+1), k(t+1), \Omega)}{\lambda} \right) + \lambda\zeta}_{Z(s(t+1), k(t+1), \Omega)} \mid s(t), k(t), d_j(t) \right], \end{aligned}$$

where λ is the scale parameter of the type-1 extreme value distribution governing the preference shocks and ζ is the Euler constant. The expectation in the last two equalities is over the type transition. Given our assumptions on the type transition process, we get

$$\begin{aligned} &\mathbb{E}\big[V(s(t+1), k(t+1), \Omega) \mid s(t), k(t), d_j(t)\big] \\ = \;& \mathbb{E}_{k(t+1)}\big[Z(s(t+1), k(t+1), \Omega) \mid s(t), k(t), d_j(t)\big] \\ = \;& \big(1 - p(t+1)\big)\big(Z(s(t+1), k(t), \Omega) + p(t+1) \sum_{k=1}^{K} \big(q_k(t+1) Z(s(t+1), k, \Omega)\big)\,. \end{aligned}$$

With the value functions properly defined, the model is solved with backward induction. At the last period, T the individual chooses $d_j(T)$ that maximizes his utility, then at period

$T-1$ he can compute the expected value function for each possible point in the state space and choose the decision that maximizes his value function in $T-1$ and so on until the first period. After solving the dynamic programming problem, one obtains the expected value functions for all possible state points.

3.4.2 Maximum Likelihood Estimation

We estimate the model using Maximum Likelihood. Each individual's likelihood is made of three components: choices, wage, and attrition.

Given preference shocks, $\varepsilon_j(t)$, have a type-1 extreme value distribution, the probability that an individual of type $k(t)$ chooses alternative j at age t is given by

$$Pr\Big(d_j(t)\Big|s(t),k(t),\Omega\Big) = \frac{\exp\Big(\widetilde{V}_j(s(t),k(t),\Omega)/\lambda\Big)}{\sum_{l=1}^{3}\exp\Big(\widetilde{V}_l(s(t),k(t),\Omega)/\lambda\Big)}\,.$$

We assume that observed log wages, $\log W(t)$, are measured with an error that is normally distributed, with mean 0 and variance σ_w^2. This gives the wage density

$$Pr\Big(\log w(t)|s(k),k(t),\Omega\Big) = \phi\left((\log W(t) - \log w(t))/\sigma_w^2\right)\,,$$

where $\phi(\cdot)$ is a density of a standard Normal distribution. The wage density appears in the likelihood when the individual decides to work in period t.

Lastly, the probability to leave the sample at a given age at which an attrition event happens takes a logit form given the type-1 extreme value distribution of the attrition shock, $\varepsilon_a(t)$:

$$\Pr(a(t)=1\,|\,s(t),k(t),\Omega) = \frac{\exp(a(t)^*)}{1+\exp(a(t)^*)}\,.$$

This probability appears in the likelihood at most three times (at the three survey years) in an individual's sequence of decisions.

The individual type-specific contribution to the likelihood at age t then takes the following form:

$$
\begin{aligned}
L(\Omega, D(t), W(t), A(t)|s(t), k(t)) \quad = \quad & \Pr(a(t) = 1 \,|\, s(t), k(t), \Omega)^{A(t)} \times \\
& \Big[(1 - \Pr(a(t) = 1 \,|\, s(t), k(t), \Omega)) \times \\
& \Pr\left(d_1(t) = 1|s(t), k(t), \Omega\right)^{D_1(t)} \times \\
& \left\{ Pr\left(d_2(t) = 1|s(t), k(t), \Omega\right) \cdot \Pr\left(w(t)|s(t), k(t), \Omega\right) \right\}^{D_2(t)} \times \\
& \Pr\left(d_3(t) = 1|s(t), k(t), \Omega\right)^{D_3(t)} \Big]^{(1-A(t))},
\end{aligned}
$$

where $D_j(t)$ is a dummy variable equal to 1 if observed choice j is taken at period t, 0 otherwise and $D(t) = \{D_1(t), D_2(t), D_3(t)\}$ and $A(t)$ is a dummy variable equal to 1 if the individual is observed leaving the sample at period t, and 0 if he is observed staying.[8]

The individual likelihood of any given path of decisions, wages, and attrition is derived by forming the product of age and type-specific likelihoods $L(D(t), W(t), A(t), \Omega|s(t), k(t))$ from age 16 to age T_i, where T_i is the maximum age until which individual i is observed, and then integrating it over the unobserved heterogeneity distribution:

$$
\begin{aligned}
L_i(\Omega, D, W, A|s(16)) = \sum_{k_{16}=1}^{K} \sum_{k_{17}=1}^{K} \cdots \sum_{k_{T_i}=1}^{K} \Big[& q_k(16) L(\Omega, D(16), W(16), A(16)|s(16), k(16)) \times \\
\prod_{t=17}^{T_i} \Big(& (1 - p(t)) L(\Omega, D(t), W(t), A(t)|s(t), k(t) = k(t-1)) + \\
& p(t) \sum_{l=1}^{K} q_l(t) L(\Omega, D(t), W(t), A(t)|s(t), k(t) = l) \Big) \Big],
\end{aligned}
$$

where D, W, and A are the vectors of observed choices, wages, and attrition events over ages. The model is estimated with $K = 2$ types.[9]

3.5 Results

We present two sets of preliminary results with two types of unobserved heterogeneity. The first set estimates the time-unvarying model with and without attrition correction. The second set of results gives preliminary evidence of time-varying components in the model's unobserved heterogeneity, without accounting for attrition.

[8] For ages at which there is no attrition event, the contribution to the likelihood only includes the choice and wage parts.

[9] In future work, we will estimate the model with a larger number of types and choose the optimal number of types by comparing information criteria, e.g., BIC or AIC.

he first set of results is obtained by estimating two types time-unvarying unobserved heterogeneity accounting and not for attrition models. They are shown in Table 3.5. The results correcting or not for attrition obtained are extremely close. An extra year of schooling raises wages by 2%, but an extra year of work experience raises wages by nearly 7%. Being at school after the age of 18 has a negative effect, while the value of being at home decreases with age. Type 1 gets lesser rewards in each choice than type 2 in the unobserved types. Type 1 students, for example, receive 53,643 euros per year to attend school, whereas type 2 students are rewarded with 95,459 euros. Similar differences occur between both types in the rewards of staying at home and working. The unobserved heterogeneity distribution for type 1 shows that having an executive or white-collar mother and father lowers the probability of being type 1 by 0.179 percentage points, although being late in school raises it. In both models, an individual has around 72 percent chance of being type 1 and close to 28 percent chance of being type 2.

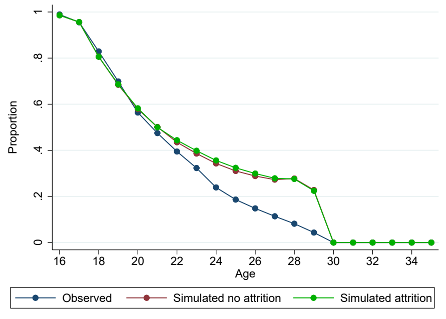
Correcting for attrition does not change the estimated parameters related to the choices or the unobserved heterogeneity distribution. But the attrition equation parameters show that years of schooling and work experience increase the probability that an individual leaves the sample. Not controlling for attritions means that low-educated individuals are more likely to leave the sample earlier, creating an upward bias in the way wages evolve with experience (as more educated individuals would be over-represented in the sample with large years of experience). This would result in an overestimation of schooling attainments.

Table 3.5: Parameters estimates of two types time-unvarying unobserved heterogeneity models correcting or not for attrition

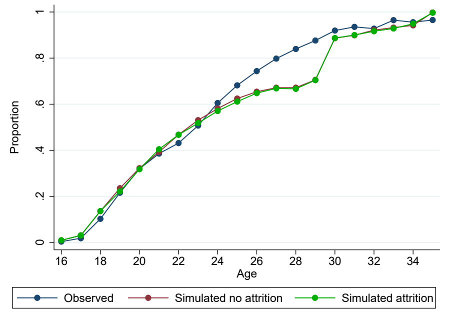
	Attrition		No attrition	
	Parameter	S.D	Parameter	S.D
1. Work				
Urban	0.016	0.002	0.016	0.002
School years	0.021	0.001	0.021	0.000
Experience	0.069	0.001	0.069	0.001
Experience ≤2	0.109	0.004	0.110	0.004
Intercept - Type 1	2.028	0.010	2.027	0.010
Intercept - Type 2	2.474	0.012	2.473	0.012
Log(standard error)	-1.628	0.005	-1.628	0.004
2. School				
Urban	2.318	0.197	2.320	0.167
Age ≥18	-21.061	0.897	-21.016	0.799
Intercept - Type 1	53.643	1.196	53.593	1.157
Intercept - Type 2	95.459	1.721	95.324	1.678
3. Home				
Age	-0.999	0.047	-1.002	0.045
Intercept - Type 1	62.404	1.688	62.517	1.629
Intercept - Type 2	96.822	2.235	96.812	2.140
4. Primitives				
Std of preference shock	11.620	0.352	11.590	0.322
5. Unobserved heterogeneity distribution of Type 1				
Both parents French	0.134	0.094	0.137	0.101
One parent French	-0.015	0.130	-0.012	0.139
Late School	1.183	0.087	1.201	0.092
Technical or Blue Collar father	0.263	0.077	0.268	0.090
Executive or White Collar father	-0.318	0.067	-0.318	0.078
Technical or Blue Collar mother	0.171	0.090	0.172	0.111
Executive or White Collar mother	-0.179	0.064	-0.183	0.080
Intercept	0.753	0.105	0.758	0.112
6. Attrition				
Years schooling	-0.058	0.010		
Experience	-0.050	0.012		
Couple	0.019	0.050		
One kid	-0.342	0.086		
Two kids	-0.182	0.128		
Urban	0.017	0.055		
Distance from MS	0.031	0.013		
Intercept - Type 1	0.183	0.149		
Intercept - Type 2	0.417	0.167		

Figure 3.2 shows the fit of both models with respect to the observed data. Figure 3.2e shows the fit in the probability to leave the sample. As mentioned above, correcting for attrition does not change the estimates of the model, and hence, the fit of both models is almost the same.

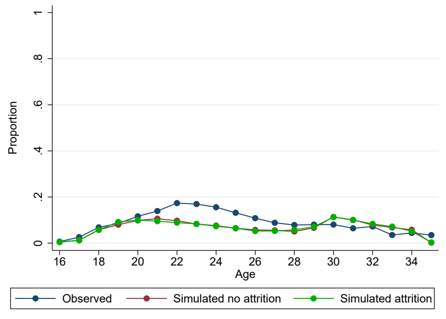
However, simulations exercise can differ as the state variables of the model affect the attrition probability.



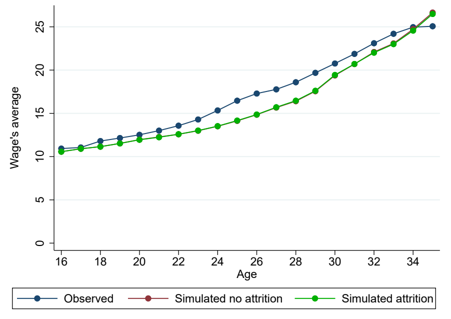
(a) School.



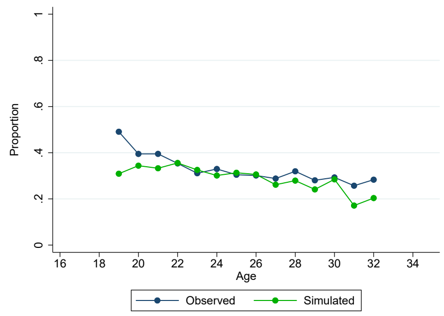
(b) Work.



(c) Home.



(d) Wages.



(e) Attrition.

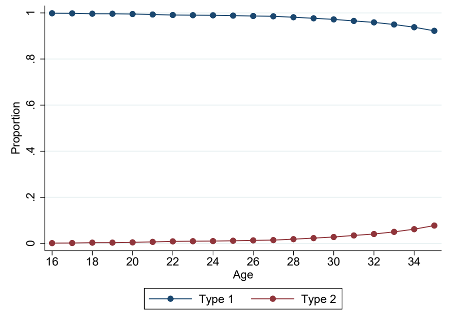
Figure 3.2: Fitting of the time-unvarying models

The second preliminary set of findings focuses on the time-varying component of unobserved heterogeneity. For the time being, we are concentrating on a dynamic model without attrition correction. Our model's time-varying unobserved heterogeneity is constructed up of two parts. The first is the probability that individuals change types throughout their age, and the second is the probability of each type changing across time. Table 3.6 presents the estimation results of the time-varying components of the model, estimated by fixing the utilities' related variables in Table 3.5 (variables in panels 1, 2, 3, and 4). By allowing the unobserved heterogeneity distribution for type 1 to depend on the accumulated years of schooling, it changes the rest of the parameters such that the probability is almost 1, making most of the sample type 1 for the initial ages, as shown in Figure 3.3a. However, accumulating schooling years decreases the probability of being type 1 by 0.5 for every additional year. This change makes more educated individuals receive higher returns later in life as the rewards to type 2 are larger for every choice than for type 1. For example, an individual at age 26 with 15 years of schooling, and 5 years of experience will gain close to 15,000 euros per year if he's type 1, while if he changes to type 2, his wage will be around 23,500 euros per year.

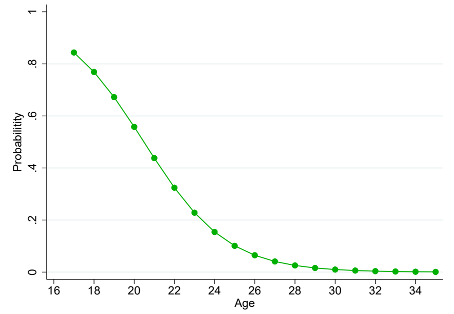
Table 3.6: Parameters estimates of two types time-varying unobserved heterogeneity model not correcting for attrition

	Parameter	S.D
1. Unobserved heterogeneity distribution of Type 1		
Both parents French	-0.245	0.103
One parent French	0.017	0.132
Late School	5.258	0.687
Technical or Blue Collar father	0.740	0.091
Executive or White Collar father	-0.175	0.064
Technical or Blue Collar mother	-0.267	0.107
Executive or White Collar mother	-0.512	0.073
Intercept	12.269	0.237
School years	-0.503	0.012
2. Probability to change types		
Intercept	-2.170	0.210
Age	0.484	0.020

Related to the probability to change types, it behaves in a convex way as shown in Figure 3.3b. Based on our estimates, the probability of changing type is around 0.84 (percentage points) at the age of 17 and then diminishes to 0.01 (percentage points) around age 30. In other words, the older the individual, the less likely he is to change types, becoming relatively fixed by the time an individual reaches age 30. This increases in type's persistence with age is in line with the one obtained by Todd and Zhang (2020).



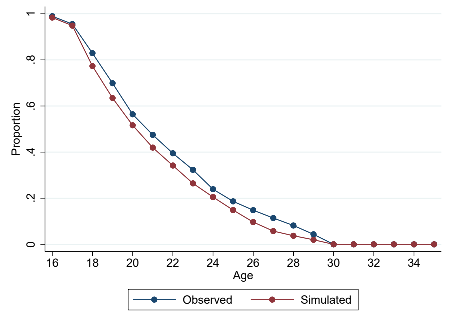
(a) Proportion of types across ages.



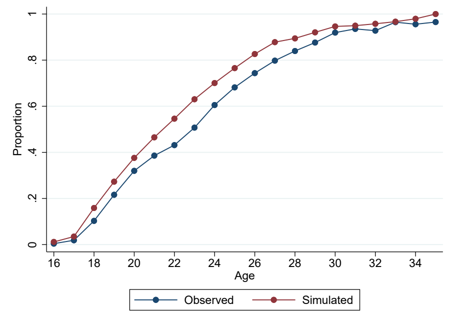
(b) Probability to change types across ages.

Figure 3.3: Time-varying unobserved heterogeneity

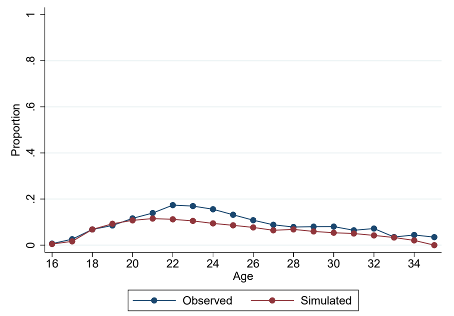
Related to the fit of the time-varying model, it improves with respect to the time-unvarying case except for the wages. The model underestimates the wages at later ages, which is generated by the low number of type 2 individuals, which are the ones with higher rewards.



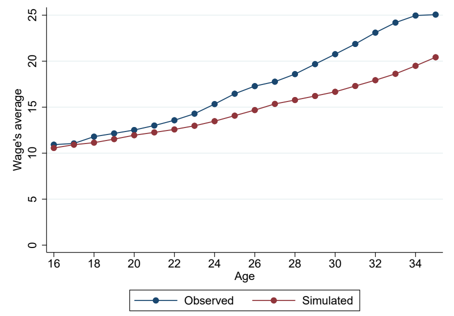
(a) School.



(b) Work.



(c) Home.



(d) Wages.

Figure 3.4: Fitting of the time-varying model

3.6 Conclusions

Dynamic structural models of human capital accumulation have been used by economists for modeling individuals' life-cycle decisions and providing relevant policy recommendations. The incorporation of time-unvarying unobserved heterogeneity is an important component of these models. Unobserved characteristics, which include cognitive and non-cognitive, have a direct impact on educational attainment and labor market success. An increasing body of empirical evidence indicates that these abilities change with age and can be influenced by education. Another aspect that has gone ignored is the presence of missing observations due to attrition. If the reasons behind attrition are endogenous, i.e., related to individual characteristics that influence schooling choices and educational outcomes, or directly related to the choices and outcomes, the absence of these observations would result in a selection bias in the estimations.

We show in this paper that these two essential methodological issues are to be considered while estimating these models. First, for attrition, our preliminary results show including unobserved heterogeneity in the model can adjust for potential attrition bias, attrition can be correlated with state space variables, affecting simulations used in policy recommendations. In our estimation, attrition is affected by years of schooling and work experience, making it most likely that highly educated and with more work experience to stay longer in the data, hence making long-term simulations adequate for these individuals only.

The second methodological issue we address is the time-unvarying assumption. We relax this assumption by allowing individuals to change types over time, as well as the probability of each type changing over time. Differing from previous empirical exercises, we only use state space variables to allow the time-varying component in the transition probability. Preliminary results indicate a time persistence difference in the transition probability of unobserved heterogeneity. Schooling increases the probability of being of the type with larger rewards for later-life decisions. This is a new indirect effect of individual human capital accumulation. We found that the probability of people changing types decreases with age, approaching zero at the age of 30 as in Todd and Zhang (2020).

The next step is to estimate the time-varying unobserved heterogeneity model to examine how the types' rewards and utilities change. Because of a high correlation between the age-related variables due to how the data was created, we might use exogenous time-varying variables such as the number of children in the unobserved heterogeneity distribution. Then, we will include the attrition correction method. Lastly, we will estimate the counterfactual of increasing tuition fees, which reduces the number of years of schooling for individuals. The reduction in accumulated years of schooling has a direct influence on salaries since it reduces an individual's accumulated human capital, but from our preliminary results, it will also have an indirect effect through changes in the unobserved heterogeneity types. Because cumulative years of schooling impact the probability of attrition, it will also change the sample of individuals in the simulations.

Bibliography

Alfo, M. and Maruotti, A. (2009). A selection model for longitudinal binary responses subject to non-ignorable attrition. *Statistics in medicine*, 28(19):2435–2450.

Arcidiacono, P. (2005). Affirmative action in higher education: How do admission and financial aid rules affect future earnings? *Econometrica*, 73(5):1477–1524.

Arcidiacono, P. and Miller, R. A. (2011). Conditional choice probability estimation of dynamic discrete choice models with unobserved heterogeneity. *Econometrica*, 79(6):1823–1867.

Beffy, M., Fougere, D., and Maurel, A. (2012). Choosing the field of study in postsecondary education: Do expected earnings matter? *Review of Economics and Statistics*, 94(1):334–347.

Belzil, C. and Poinas, F. (2018). Estimating a model of qualitative and quantitative education choices in france.

Cunha, F., Heckman, J. J., and Schennach, S. M. (2010). Estimating the technology of cognitive and noncognitive skill formation. *Econometrica*, 78(3):883–931.

Ding, W. and Lehrer, S. F. (2014). Understanding the role of time-varying unobserved ability heterogeneity in education production. *Economics of Education Review*, 40:55–75.

Gayle, G.-L. and Viauroux, C. (2007). Root-n consistent semiparametric estimators of a dynamic panel-sample-selection model. *Journal of Econometrics*, 141(1):179–212.

Hu, Y. and Shum, M. (2012). Nonparametric identification of dynamic models with unobserved state variables. *Journal of Econometrics*, 171(1):32–44.

Keane, M. P. and Wolpin, K. I. (1997). The career decisions of young men. *Journal of political Economy*, 105(3):473–522.

Kyriazidou, E. (2001). Estimation of dynamic panel data sample selection models. *The Review of Economic Studies*, 68(3):543–572.

Little, R. J. and Rubin, D. B. (2019). *Statistical analysis with missing data*, volume 793. Wiley.

Moffit, R., Fitzgerald, J., and Gottschalk, P. (1999). Sample attrition in panel data: The role of selection on observables. *Annales d'Economie et de Statistique*, pages 129–152.

National Longitudinal Surveys (2023). Retention rates in the NLSY97. Accessed: 22-10-2019.

Rothstein, D. S., Carr, D., and Cooksey, E. (2018). Cohort profile: the national longitudinal survey of youth 1979 (nlsy79). *International journal of epidemiology*, 48(1):22–22e.

Rust, J. (1987). Optimal replacement of gmc bus engines: An empirical model of harold zurcher. *Econometrica: Journal of the Econometric Society*, pages 999–1033.

Spagnoli, A., Marino, M. F., and Alfò, M. (2018). A bidimensional finite mixture model for longitudinal data subject to dropout. *Statistics in medicine*, 37(20):2998–3011.

Todd, P. E. and Zhang, W. (2020). A dynamic model of personality, schooling, and occupational choice. *Quantitative Economics*, 11(1):231–275.

3.A Data construction

We clean the data according to each different sub-data presented in *Génération* 98 Data: general information, working spells, and unemployed spells.

3.A.1 General data

We kept the most relevant variables to create those variables that change across time: family characteristics, attrition, and distance.

3.A.2 Working data

We kept each data separately, leaving only the relevant information regarding spells. We then merge one by one to construct the monthly wages as explained below. Each merge is done sequentially. After each, we cleaned the data to leave the most updated information. This includes creating a unique variable for the beginning, end, duration, and type of spell. This is done in 2 steps: (1) from spells, construct monthly observations, and (2) aggregate monthly observations to construct yearly observations.

Step 1: construct monthly observations. For each employment spell, we have the information on wages and working time at 2 times: the start of the spell and the end of the spell. We are using this information to extrapolate and create observations every month within the spell. Regarding the timing, we define a year from July to July, and we then extrapolate the non-seen months in the sequence. We define individuals as working each month when they work full-time (FT). They are coded as staying at home when they do not work or work part-time (PT).

- Case 1: an individual works FT at the start and FT at the end of the spell: we code the individual as working for all the months within the spell and we calculate the monthly wage using starting and ending wages.
- Case 2: an individual works PT at the start and PT at the end of the spell: we code the individual as staying at home for all the months within the spell.
- Case 3: an individual works FT at the start and PT (but at least 50%) at the end of the spell: we code the individual as working for all the months within the spell and we do not take the wage (missing value for all the months of the spell).
- Case 4: individual works PT (but at least 50%) at the start and FT at the end of the spell: we code the individual as working for all the months within the spell and we do not take the wage (missing value for all the months of the spell).

If there are missing values in full vs time work we apply the following rules:

- Drop observations in all data sets with category “sco”, “afa” or “slo” in typeseq.
- “vac” are considered unemployed.
- “asc” are fully employed.

When we merge, we consider two cases: (1) unmatch observations where we take information from each data. And (2) match observations where we are careful of two situations. As some individuals are interviewed at some moment before the end of our setting of time, we must fill up their information during the months they were not interviewed during the same wave. This procedure is only for those sequences that are not observed in the next wave of interviews. If the observation was censored at the end of the previous wave, then we take the initial information from the oldest wave and the final information from the latest wave. If instead, the observation is censored at both we take the initial and final information from the newest wave.

To calculate wages, first notice that for individuals who are observed in 2 consecutive waves, we have 3 points of observation for the wage: the wage at the start of the spell, the wage at the end of the spell (observed in the second wave) and wage at the time of the first survey. We use these 3 points of observation to extrapolate the wage (with different slopes before and after the first survey date). To do this, we go sequentially through the different databases. We replace the initial wage with the last wage from the previous data. We do the same with the initial wage to have only the number of months for each new observation of wages. Then, we follow the next steps:

1. Calculating FT-equivalent wages at the start or end of the spell (when necessary):
2. If the person works X% of the time (where X=50, 60 or 80). Take the wage, divide it by X, and multiply by 100.
3. Calculating wage for each month in a spell using the starting and ending wage: assume that the wage evolves linearly from the start to the end of the spell. Growth rate = (end wage – start wage) / spell duration in months. Wage at month t within the spell (t=0 is the starting month) = start wage + (growth rate * t)

Step 2: we aggregate monthly observations into yearly observations. We first define when to start a year of observation in July 1998. Then, for each year, we adopt the following rule to aggregate information:

- Choices:

- All the years before 1998 are assigned to school.
- If the number of months worked (according to the criterion used in step 1) is larger or equal to 6 months the choice is work.
- Otherwise, the individual stays at home.

- Wages:
  - For a year classified as Work, denote M the number of months worked in the year (obviously, M is larger or equal to 6).
  - Make the sum of the wages observed on all M months (except if we do not take the wages into account. See step 1).
  - If M is lower than 12, we extrapolate the wages to get an annual FT-equivalent wage, by taking the sum of wages, dividing by M, and multiplying by 12.

Finally, we remove inflation by transforming wages to have wages in 2001 euros. We use data from *L'Institut national de la statistique et des études économiques*'s website.

3.A.3 Unemployed data

From this data, we need the information for when was the last month the individual was seen. To obtain it, I focus if the individual was censored or not and extract the beginning date, the end date, the duration, and the number of sequences.

3.A.4 Panel

Append spells

The first step is to append the working and unemployment spells. This allows having the complete sequence for all the individuals, allowing to update the data for the individuals who are present in the different waves. With this information, we create the monthly information. It is important to mention that those sequences that are censored are filled up for the months that are censored, completing the whole month for each wave. With the monthly information created, we defined the years and create the year variables. Finally, we save two different data sets. One with all the sequences called "spells_complete". The second one, "spell_panels" has only one observation for everyone with all the information needed for the panel.

Panel

The creation of the panel goes with the general information data and the spell data set "spells_panel". We start by dropping women and individuals under age 16. Afterward, we create the choice variable and reshape the data to a long format. Finally, we create the state space variables and control variables.