Three Essays about Enforcement, Labor Markets and Education

Rodrigo Ceni Gonzalez

Thesis submitted for assessment with a view to obtaining the degree of Doctor of Economics of the European University Institute

Florence, November 2014

European University Institute
Department of Economics

Three Essays about Enforcement, Labor Markets and Education

Rodrigo Ceni Gonzalez

Thesis submitted for assessment with a view to obtaining the degree of Doctor of Economics of the European University Institute

Examining Board

Prof. Russell Cooper, Penn State University, Supervisor
Prof. Jérôme Adda, EUI & Bocconi University
Prof. Giovanni Pica, University of Salerno
Prof. Marco Manacorda, London School of Economics

© Ceni Gonzalez, 2014
No part of this thesis may be copied, reproduced or transmitted without prior permission of the author

Abstract

This thesis analyzes how government enforcement contribute to the labor market and educational behavior in developing countries. The first chapter studies how informality responds to the quality of the labor enforcement and the bundle of benefits that the formal workers receive. Countries in Latin America with different levels of informality were compared, highlighting the features that could induce these different levels. In a general equilibrium framework, the government chooses a level of enforcement and a bundle of benefits maximizing the workers utility subject to a budget constraint. A representative firm chooses the share of workers in formality and informality that they want to hire, and the workers offer a share of time in formality and informality. The chapter concludes that differences in the quality functions of government enforcement and benefits are found, as well as in the fines established to enforce the agents. The second chapter, co-authored with Gonzalo Salas, examines how the level of enforcement of the conditionalities of two Conditional Cash Transfer programs affects the ratios of high school students drop-out. We develop a structural discrete choice model in which the individuals who are above or below the participation threshold decide whether or not to attend school, participate in the labor market, or spend time on home production and/or leisure. The policy experiments show that if the level of enforcement is higher, individuals change study for leisure and work, but this last choice has a limit. Moreover, if the amount of transfer is reduced, the share of those who only study goes down and individuals work more. The third chapter examines how changes in the social security scheme affect the participation path of workers between formality and informality. Workers construct their decision paths in the labor market depending on the retirement program and their endowment of human capital. The strictness of the requirements lead to more formality but not enough to obtain a pension for all the educative levels. Finally, the extension of the compulsory active life leads to more formality and better pensions.

To Luisa

Acknowledgements

Foremost to Luisa for her love, support, patience and encouragement. To Isabella and Emilio for giving to this work the right balance in my life. My parents, sisters and grandparents for their love. Tommaso, Jenifer, Vasco and Agostina for being our family in Italy. I would like to express my sincere gratitude to my supervisors Russell Cooper and Jerome Adda for their patient guidance, continuos support and valuable critiques. A special thank goes to Andrea Ichino who supervised my work at the beginning of the PhD and offered me genuine help. My deep gratitude goes to my friend and coauthor Gonzalo Salas. My thanks to my *Florentinean* friends Silvia, Stefan, Marek, Simon, Helia, Chara, David, Fede and all the mates who passed by the Statistics Room at Badia in 2009-2010, and my *Napolitanean* friends Bruno, Francesca, Ceci, Gianluca, Rosario, Andrea and Simona. I would like to acknowledge, for their help and kindness, the staff of Villa San Paolo at the EUI, Thomas Bourke, Giuseppe Feola, Marcia Gastaldo, Martin Legner, Loredana Nunni, Sonia Sirigu, Jessica Spataro, Julia Valerio and Lucia Vigna; and the staff members of The Creche.

Table of Contents

List of Figures

List of Tables

Chapter 1

Informality and government enforcement in Latin America

1.1 Introduction

Latin America is one of the regions where informality has been identified in depth. About 50% of salaried workers[1] are employed informally, if we define informal workers as those who are not covered by labor regulation, such as taxation, the right to the health system and the right to a pension income in retirement (Portes et al. (1989) and Schneider (2012)). Informal work defines an underdevelopment phase of these economies and can be better understood as multicausal and complex, and in which the government can play a crucial role. Although informality is present in each country of the region and is also present among different levels of education, as is shown in Table 1.1, the figures are extremely heterogeneous between different countries.

The role of the government in informality can be conceived in three main dimensions: first, the administration receives taxes and contributions from the formal workers; second, the regulator monitors the firms looking for informal jobs, fines them and eliminates these jobs; and third, the administration brings benefits to the formal workers through the health system, pensions and unemployment payments. These three assignments are carried out to different degrees and lead to different levels of enforcement for firms and workers. **The aim of this paper is to shed light on how the labor market responds to the quality of the government enforcement and the quality of the bundle of benefits brought by the public institutions.** To reach this objective I compare the informality performance in different countries, given the quality of the enforcement of the public sector. **I develop a model to capture these relationships and I estimate it for five countries with different levels of informality in order to capture the heterogeneity.** These five countries are Argentina, Brazil, Colombia, Peru and Uruguay.

In this paper I will only measure the informality of those who declare that their em-

[1] Salaried workers are those who have a salary paid by a employer.

ployer does not pay the necessary contributions in order to give them the right to a pension in old age. The decision to use this definition is due to the fact that this question is present in all the household surveys and its consequences have been widely analyzed in the literature (Holzmann and Takayama (2009), Joubert (2012) and the Chapter 3 of this Thesis). Levels of informality are clearly heterogenous among countries, not only at all the educative levels[2] but also where other benefits are considered in the definition of informality, such as health benefits or the 13th salary as is shown in Table 1.2.

Country	Mean	Low education	High education
Argentina	36.1%	39.7%	15.2%
Brazil	21.6%	26.8%	14.3%
Colombia	42.7%	46.9%	14.8%
Peru	51.6%	66.8%	25.2%
Uruguay	19.4%	25.0%	6.5%

Table 1.1: Informality rates among salaried workers in 2009. Source: CEDLAS, World Bank.

Analyzing the nature of the informality, there are three strands in the literature. The first one has extensively claimed that there are two separate segmented markets, which have different rules, and have been related with low and high productivity sectors. From the firm point of view, formal and informal workers are treated as two different inputs. This concept has been discussed by the empirical literature using data from Mexico, Colombia, Argentina and Uruguay[3], where the evidence suggests that it is the workers who decide whether to be formal or informal employees. The second strand focuses on the workers' individual decision to be in each sector given their characteristics as a unique labor market. Finally, the third one proposes a moderate dualism, which is considered in most of the recent theoretical papers. Workers and firms can decide to operate optimally either formally or informally. Galiani and Weinschelbaum (2007) present a model of an economy with a continuum of firms and workers in which formal and informal jobs and their wages are endogenous. The main feature which determines formality is the managerial ability which drives the difference in human capital between these two sectors.

Amaral and Quintin (2006) present a dynamic model where managers can either self-

[2]High educated workers are defined as those who at least finished high school, and the low educated the one that did not.

[3]Magnac (1991), Maloney (2004), Pratap and Quintin (2006) and Bucheli and Ceni (2010)

finance part of their capital with savings or borrow funds from an intermediary. In this way, the most talented managers self-select into the formal sector in which formal managers operate with more physical capital than informal managers and informal employers self-finance more intensely than the formal ones. The model reproduces the main macro elements of labor markets in developing nations.

The main controversial point in these models is that the worker's decision does not play any role in the equilibrium, which is in contrast to the literature which largely accepts that it does, especially in the middle and high educated workers. The central argument against this dualism is the mobility between these sectors, which is clearly observed in these economies, if workers move between sectors it means that their intrinsic characteristics are not so different, then it can be considered as the same input in the production function. Mobility is a permanent feature of this phenomenon (Fields (2011)). Indeed, there were a significant number of annual changes among sectors in Argentina between 2003 and 2011 (Table 1.3)

My model introduces a novel general equilibrium framework, where the three main characters in the economy: households, firms and government optimize the level of informality, government enforcement and the benefits that formal workers receive. Additionally, **I allow a loose definition of dualism through the definition of the production function of the representative firm which includes formal and informal workers, where the level of substitution is specifically estimated**. In contrast with the main papers in the literature, which focus on entrepreneurial ability and how it determines in which sector they develop their activity; firms hire both formal and informal workers contemporaneously, and enter in the production function with different levels of substitutability by education. If both inputs are perfect substitutes, the optimal is the corner solution and inputs can be treated as only one. But, if the level of substitutability is lower, both inputs coexist in the production function. Theoretically this is one of the main contributions of this paper.

There are cost and benefits for those either in formality or informality, and these are borne by employees and employers. Firstly, in formality the employees are obliged to pay contributions and taxes, but they have the right to receive benefits in the present and in the future (such as the right to be covered by the health system, enjoy holidays, receive some extra payments and a pension for the elderly). However, some workers do not value some of these benefits, because the services which are provided are of poor quality, or the

Country	Pensions	Health	13th month	Holidays
Argentina	36.1%	35.9%	34.6%	34.4%
Brazil	21.6%	-	-	-
Colombia	42.7%	42.4%	53.7%	28.2%
Peru	51.6%	51.3%	-	-
Uruguay	19.4%	21.5%	25.3%	-

Table 1.2: Informality levels using different benefits in 2009. Source: CEDLAS.

Probability of yearly change of sector			
	Unemployment	Formal	Informal
Unemployment (-1)	0.337	0.254	0.409
Formal (-1)	0.024	0.909	0.067
Informal (-1)	0.086	0.240	0.674

Table 1.3: Probability of change of sector in Argentina, based on the multinomial model 2003-2010. (only men)

government commitment is too weak *i.e.* they believe that in the future a form of survival pension will be available for everyone. Additionally, informality is attractive for some workers because it is a more flexible sector, which allows easier entrance for unemployed workers or for those who want to acquire experience without signing a formal contract.

The World-Bank (2010) also focused on governmental policies designed to improve the life quality of the poorest population, such as health insurance or Conditional Cash Transfer programs. In the case of Colombia, the government has a program to provide health services to the population below a formal income threshold, so the workers prefer to be informal because there are no additional (or there are lower) benefits to working formally.

Secondly, firms in informality do not pay any contributions and taxes because they are not monitored by the government or the fines are too low. Moreover, there is no social punishment because this is not viewed as a crime. However, in formality they benefit from government protection against possible abuse from criminal activities, such as blackmail (Loayza et al. (2009)).

Finally, the role of government emerges as crucial in both academic and political discussions about informality. The weakness and corruption of governments play a negative role in the analysis. Conversely, there is a trade-off about the burden of taxes and reg-

ulations: on one hand stricter laws can dissuade firms and workers from formality, but, on the other hand stricter regulation and monitoring should prevent informality (Loayza et al. (2009)). Moreover, informality plays a role in these economies as a source of flexibility in the cyclical phase of recessions by absorbing the unemployment which could be generated by the impossibility of paying the cost of the formality. This counter cyclical behavior leads to governments allowing this practice in order to moderate the impact of the economic downturn (Loayza and Rigolini (2011)).

Ihrig and Moe (2004) analyze the size of the shadow economies with a two sector model when the taxes and the enforcement of the taxes change. The main objective of their paper is to explore the size and the dynamic of the shadow economy cross-country, and when the enforcement of the taxation changes exogenously. The definition of shadow economies is more reductive than that of informality because it only focuses on the taxation side of the problem. Moreover, the government role is exogenous in contrast with my paper when the government choices are endogenous.

The research question that guide this chapter is: What are the **underlying differences** (e.g. in enforcement technology, the ability to provide public benefits, the production technology, the skill composition of the labor force) that explain jointly the **choice of governments on the level of government enforcement and benefits, and the reaction of firms and workers in terms of operating in formality/informality**? My aim is to compare the informality performance in different countries given the quality of the bundle of benefits, the costs, fines and quality of the government enforcement.

The main objective is to explain the heterogeneity, regarding the fact that this phenomenon is present in all countries in Latin America, and **explore the features that could explain these differences, focusing on costs, benefits and enforcement**. I estimate the main parameters which characterize each country (Argentina, Brazil, Colombia, Peru and Uruguay) in order to assess the differences in the informality levels.

This paper is structured as follows. In Section 2 I introduce the data and the main variables of the paper, Section 3 provides the model, with one period and two types of workers with different educative levels, Section 4 presents the main results of the estimation and the experiments, and finally, in Section 5 I present the main conclusions.

1.2 Data and main variables

I use data from five countries with different levels of informality: Argentina, Brazil Colombia, Peru and Uruguay. In order to estimate the production function I use data from the National Accounts and the household surveys for each one. The data from the National Accounts is the Gross Value Added for seven sectors (primary, manufacturing industries, construction, commerce, transport, finance and services).The household surveys have a socioeconomic purpose and they are crucial in identifying workers in different productive sectors in the economy. Identifying the formal workers[4] is done directly by asking if the employer pays the contribution in order to obtain the right for a pension in retirement. The high educated workers are identified as those who declare that they completed high school (completed high school and higher), and the low educated all the other ones (uncompleted high school and lower).

I estimate the informality by education and sector quarterly, and the GDP for each sector quarterly. In this way the data-base to estimate the production function has 1,162 observations (414 for Argentina, 88 for Brazil, 108 for Colombia, 240 for Peru and 312 for Uruguay).

For Argentina, I use the Permanent Household Survey (EPH in Spanish) carried out by the National Institute of Statistics and Census (INDEC in Spanish) for the period 1995-2010. The sample is restricted to the urban regions, covering 28 large urban centers where 70% of the urban population of Argentina live[5].

For Brazil, I use the Continuous Household Survey (PNAD)[6], conducted by IBGE[7] in September of each year between 1996 and 2007. The survey is only carried out in September, so I only have one observation per year.

In the case of Colombia, I use the Continuous Household Survey (ECH) between 2002 and 2005 and the Large Integrated Household Survey (GEIH) between 2007 and 2010, conducted by the National Bureau of Statistics (DANE). The question about the social contribution is only present in the second quarter in the ECH and in the first half of the

[4]Note that, informal workers are those who are not covered by labor regulation, such as taxation, the right to the health system and the right to a pension income in retirement.

[5]Urban population accounts for the 90% for the total population of Argentina, so the survey gives a good representation of the country.

[6]Pesquisa Nacional por Amostra de Domicilios.

[7]Instituto Brasileiro de Geografia e Estatística.

year in the GEIH, so the number of observations are limited.

I use the National Household Survey (ENAHO) of Peru carried out by the National Institute of Statistics and Informatics (INEI) in the period 2001-2010. The sample includes all urban and rural areas in the country.

In the case of Uruguay, I use the Continuous Household Survey (ECH) conducted by the National Statistics Institute (INE), between 1997 and 2010 throughout the whole year. The ECH is a survey carried out in urban areas between 1997 and 2005, where more than 90% of the Uruguayan population is, so the survey gives a good representation of the country. From 2006 the survey includes rural areas as well.

Finally, to estimate government enforcement and the quality of the benefits that the formal workers receive, I use some indicators collected by the InterAmerican Development Bank. In particular for the quality of government enforcement I use two indicators: compliance with the law[8] and confidence in the judiciary system[9]. For the benefits that the government provides, I use the citizens' perception of the taxes being well spent[10].

In this framework, the literature basically assigns three roles to the government: collecting taxes, providing benefits and monitoring and regulating the economy. Empirical analysis shows an ambiguous relation between the unofficial economy and the level of taxation, and a positive relation with corruption using a large cross-country data-base of entrepreneurs. Johnson et al. (1998), using simple OLS regression, find a positive relation between regulation bureaucracy, tax burden and corruption with higher unofficial activities. Friedman et al. (2000) go further with a larger number of countries and find that the taxes have a negative effect on the unofficial economy, and it is the corruption and the bureaucracy which have a positive effect.

First, in Figure 1.1 we can observe the relationship between informality and the social contributions (taxes) of both employee and employer. There is a slightly negative relation between them, i.e. countries with a high level of contribution have less informality. This relation is observed both with the total, employer and employee contribution, and it could be interpreted as a part of the institutional framework.

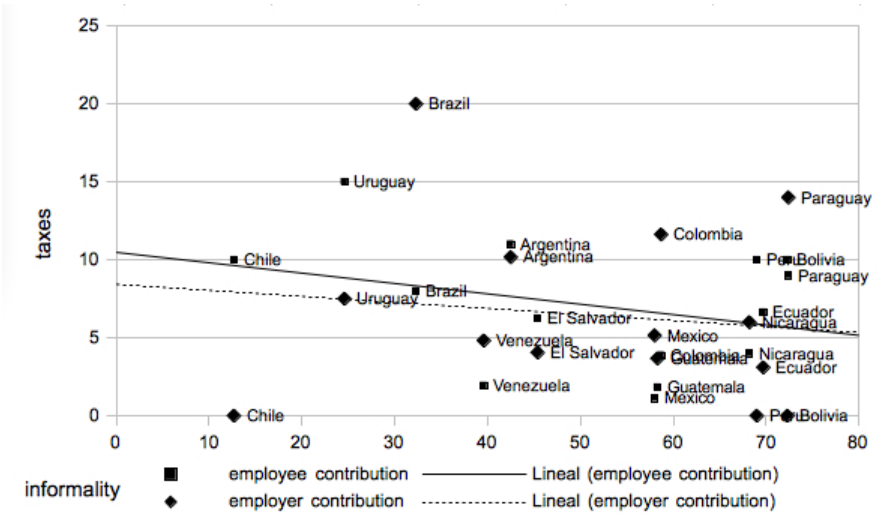
[8]This indicator represents the percentage of those surveyed who respond that they believe that citizens comply with the law very much or a fair amount. Source: Latinobarometer.

[9]It measures the percentage of firms that agree with the statement: I am confident that the judicial system will enforce my contractual and property rights in business disputes. Source: World Bank.

[10]This indicator represents the percentage of answers to the question: Do you believe that the government spends your tax dollars well?

Figure 1.1: Informality by social contribution in Latin America

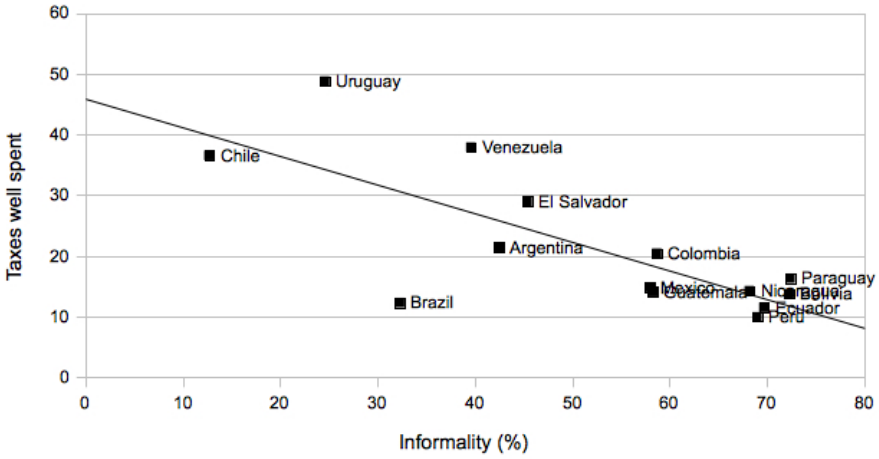
(a) Employer and employee contribution



Second, I explore the relation between informality and the quality of the benefits. In Figure 1.2 we can establish that there is a clear negative relation between the perception that taxes are well spent by the government and informality.

Figure 1.2: Informality by benefits in Latin America

(a) Taxes well spent

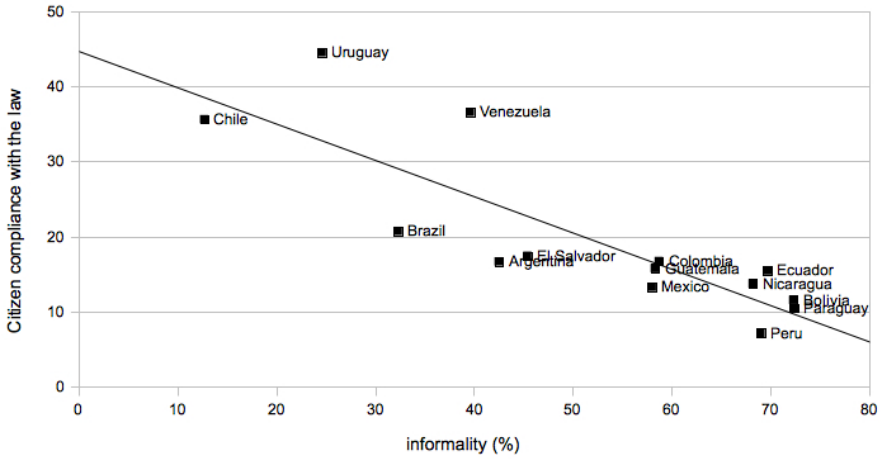


Finally, I consider the relation between informality and government enforcement, measured by citizen compliance with the law (Figure 1.3). There is a clear negative relation between them, which is in line with the literature on the relation between informality and institutions.

These three figures show that two of the three relations between government role and informality appear clearly in the cross-country data. The benefits that the workers receive and the level of government enforcement in the labor market seem to play an important role in the level of informality, and in the process of formalization.

Figure 1.3: Informality by government enforcement in Latin America

(a) Compliance with the law



1.3 Model

I develop a simple model to analyze the relation between the levels of informality by education $\left(\theta_L^i, \theta_H^i\right)$ and the quality of government enforcement $\left(q(e)\right)$ and the quality of benefits $\left(K(\kappa)\right)$. The households and firms decide on the level of informality (formality), and simultaneously there is a government choosing the level of enforcement in the labor market, and the benefits that the formal workers would enjoy.

1.3.1 Optimization problem: households and firms

Representative household problem: In a representative household, there are a continuous of x workers with a low level of education $(s = L)$, and $1 - x$ with a high level $(s = H)$. This household maximizes its consumption deciding the share of informal (θ_s^i) and formal (θ_s^f) work for each level of education s. There is no utility of leisure, the worker decides how to split their total number of hours (x and $1 - x$) between formality and informality. The consumption is determined by the revenues from formal and informal work, a lump sum tax (T) and the profit from firms (Π) which is fixed at zero. The formal

revenues are the wage of formal hours $\omega_s^f \theta_s^f$ and they also receive $\big(K(\kappa)\big)$ representing the quality of the benefits that the formal workers enjoy. The informal revenues are the informal wage $\omega_s^i \theta_s^i$ for the hours worked in informality, and there is also a share $(\phi_2 q(e))$ of this total wage which is lost. This loss depend on the quality of the enforcement $(q(e))$ and a parameter ϕ_2, and represents a market imperfection in the informal labor market.

$$\max_{C,\{\theta_s^i,\theta_s^f\}} \quad U\big(C\big) \tag{1.1}$$

Subject to

$$C = \sum_s \left[\omega_s^i \theta_s^i \Big(1 - \phi_2 q(e)\Big) + \omega_s^f \theta_s^f \Big(1 + K(\kappa)\Big)\right] - T + \Pi \tag{1.2}$$

The maximum of low skill hours is x:

$$\theta_L^i + \theta_L^f \leq x \tag{1.3}$$

and the maximum of high skill hours is $1 - x$:

$$\theta_H^i + \theta_H^f \leq 1 - x \tag{1.4}$$

Lagrangian function:

$$\begin{aligned} L = {} & U\big(C\big) + \lambda_1^{HH}\left[C - \sum_s \left[\omega_s^i \theta_s^i \Big(1 - \phi_2 q(e)\Big) - \omega_s^f \theta_s^f \Big(1 + K(\kappa)\Big)\right] + T - \Pi\right] \\ & + \lambda_2^{HH}\Big(\theta_L^i + \theta_L^f - x\Big) + \lambda_3^{HH}\Big(\theta_H^i + \theta_H^f - (1 - x)\Big) \end{aligned} \tag{1.5}$$

The interior solution to the household problem ($\theta_s^j \neq 0$) implies workers offering formal and informal hours in the labor market, then wages in informality after the market imperfection loss is equal to the formal wage plus the benefits:

$$\theta_L^i \frac{\partial L}{\partial \theta_L^i} = 0: \quad \theta_L^i \Big(- \lambda_1^{HH}\Big(\omega_L^i \Big(1 - \phi_2 q(e)\Big)\Big) + \lambda_2^{HH}\Big) = 0 \tag{1.6}$$

$$\theta_L^f \frac{\partial L}{\partial \theta^f} = 0: \quad \theta_L^f \Big(- \lambda_1^{HH} \Big(\omega_L^f \big(1 + K(\kappa)\big) \Big) + \lambda_2^{HH} \Big) = 0 \tag{1.7}$$

$$\theta_H^i \frac{\partial L}{\partial \theta_H^i} = 0: \quad \theta_H^i \Big(- \lambda_1^{HH} \Big(\omega_H^i \Big(1 - \phi_2 q(e)\Big) \Big) + \lambda_3^{HH} \Big) = 0 \tag{1.8}$$

$$\theta_H^f \frac{\partial L}{\partial \theta^f} = 0: \quad \theta_H^f \Big(- \lambda_1^{HH} \Big(\omega_H^f \big(1 + K(\kappa)\big) \Big) + \lambda_3^{HH} \Big) = 0 \tag{1.9}$$

Representative firm problem: A representative firm, decides to hire a share of informal l_s^i and formal workers l_s^f for each educative level s. The firm pays the formal workers $\omega_s^f l_s^f$ plus taxes τ, I am considering that the net wage (after taxes) and the taxes are only paid by the firm. Informal workers receive $\omega_s^i l_s^i$ but the firm faces a proportional fine $\phi_1 q(e)$ if that job is monitored.

$$\max_{l_s^i, l_s^f} \quad \Pi = \quad y\big(l_s^i, l_s^f\big) - \sum_s \Big(\big(1 + \phi_1 q(e)\big) \omega_s^i l_s^i + \omega_s^f l_s^f (1 + \tau) \Big) \tag{1.10}$$

F.O.C.

$$l_s^i: \quad \frac{\partial y(l_s^i, l_s^f)}{\partial l_s^i} - \Big(\omega_s^i \big(1 + \phi_1 q(e)\big) \Big) = 0 \tag{1.11}$$

$$l_s^f: \quad \frac{\partial y(l_s^i, l_s^f)}{\partial l_s^f} - \Big(\omega_s^f (1 + \tau) \Big) = 0 \tag{1.12}$$

The market clearing condition equalizes the share of hours in formality and informality for each level of education, that the firm demands and the worker supplies:

$$l_s^i = \theta_s^i \quad ; \quad l_s^f = \theta_s^f \tag{1.13}$$

The definition of the production function is one of the contributions of this paper. In the literature most of the papers either introduce the formality and informality as substitutes as in Ihrig and Moe (2004), or treat them as complements modeled in a Cobb-Douglass framework. The functional form which I choose is the CES function as in Dolado et al. (2001), Giuliodori and Stucchi (2010), Cappellari et al. (2011) who model the coexistence of temporary and permanent workers to reflect the fact that there are two types of workers who are not different in essence but contractually. In the informality literature, Ulyssea (2010) also presents a model with a CES production function with formal and

informal intermediate goods. The CES function allows me to introduce the loose form of market duality, if formal and informal are perfect substitutes the solution tends to be a corner solution. However, if there is an imperfect substitution, formality and informality coexist in the production function.

Remark 1 *The CES production function of the representative firm including contemporaneously formal and informal workers captures the market duality through the level of substitutability δ_j. If inputs have high substitutability, it shows that both inputs are more similar than in the case when the parameter goes to the complementarity. The production function is:*

- $y = \gamma l_H^{\rho_1} l_L^{\rho_2}$
- $l_H = \left[\psi_1 (l_H^f)^{-\delta_1} + (1-\psi_1)(l_H^i)^{-\delta_1}\right]^{\frac{-(\upsilon_1)}{\delta_1}} \quad l_L = \left[\psi_2 (l_L^f)^{-\delta_2} + (1-\psi_2)(l_L^i)^{-\delta_2}\right]^{\frac{-(\upsilon_2)}{\delta_2}}$

The level of substitutability is determined by δ_j, if it is close to -1, both inputs are perfect substitutes. Conversely if both inputs are complements, $\delta_j \to \infty$.

From the F.O.C. of the firm problem

$$\begin{aligned}
\gamma \upsilon_2 \rho_2 \theta_H^{\rho_1} \theta_L^{\rho_2 - 1} \Big(\psi_2 (\theta_L^f)^{-\delta_2} + (1-\psi_2)(\theta_L^i)^{-\delta_2}\Big)^{\frac{-(\upsilon_2+\delta_2)}{\delta_2}} (1-\psi_2)(\theta_L^i)^{-\delta_2-1} \\
= \omega_L^i \big(1 + \phi_1 q(e)\big) \\
\gamma \upsilon_2 \rho_2 \theta_H^{\rho_1} \theta_L^{\rho_2 - 1} \Big(\psi_2 (\theta_L^f)^{-\delta_2} + (1-\psi_2)(\theta_L^i)^{-\delta_2}\Big)^{\frac{-(\upsilon_2+\delta_2)}{\delta_2}} \psi_2 (\theta_L^f)^{-\delta_2-1} \\
= \omega_L^f \big(1 + \tau\big) \\
\gamma \upsilon_1 \rho_1 \theta_H^{\rho_1 - 1} \theta_L^{\rho_2} \Big(\psi_1 (\theta_H^f)^{-\delta_1} + (1-\psi_1)(\theta_H^i)^{-\delta_1}\Big)^{\frac{-(\upsilon_1+\delta_1)}{\delta_1}} (1-\psi_1)(\theta_H^i)^{-\delta_1-1} \\
= \omega_H^i \big(1 + \phi_1 q(e)\big) \\
\gamma \upsilon_1 \rho_1 \theta_H^{\rho_1 - 1} \theta_L^{\rho_2} \Big(\psi_1 (\theta_H^f)^{-\delta_1} + (1-\psi_1)(\theta_H^i)^{-\delta_1}\Big)^{\frac{-(\upsilon_1+\delta_1)}{\delta_1}} \psi_1 (\theta_H^f)^{-\delta_1-1} \\
= \omega_H^f \big(1 + \tau\big)
\end{aligned}$$

1.3.2 Equilibrium: households and firms

I will focus only on the interior solution ($\theta_s^{j*} \neq 0$) firms where both formal and informal workers coexist. The relative informal wages depend positively on the quality of benefits

and the quality of government enforcement:

$$\begin{aligned}\frac{\omega_L^{i*}}{\omega_L^{f*}} &= \frac{\big(1+K(\kappa)\big)}{\big(1-\phi_2 q(e)\big)} \\ \frac{\omega_H^{i*}}{\omega_H^{f*}} &= \frac{\big(1+K(\kappa)\big)}{\big(1-\phi_2 q(e)\big)}\end{aligned} \tag{1.14}$$

The relative size of the informal sector depends on the relative wage, the relation between fines from being in informality $\big(\phi_1 q(e)\big)$, and the contribution paid by the formal employer (τ), while the differences in educative levels is given by the formal shares (ψ_1 and ψ_2) and the level of substitutability (δ_1 and δ_2) of the production function.

$$\begin{aligned}\frac{\theta_L^{i*}}{\theta_L^{f*}} &= \left[\frac{\psi_2}{1-\psi_2}\frac{\big(1+K(\kappa)\big)}{\big(1-\phi_2 q(e)\big)}\frac{\big(1+\phi_1 q(e)\big)}{1+\tau}\right]^{\frac{-1}{\delta_2+1}} \\ \frac{\theta_H^{i*}}{\theta_H^{f*}} &= \left[\frac{\psi_1}{1-\psi_1}\frac{\big(1+K(\kappa)\big)}{\big(1-\phi_2 q(e)\big)}\frac{\big(1+\phi_1 q(e)\big)}{1+\tau}\right]^{\frac{-1}{\delta_1+1}}\end{aligned} \tag{1.15}$$

The share of informality in both educative levels in equilibrium depend negatively on the level of enforcement e (Equations 1.16) and the benefits κ (Equations 1.17).

$$\frac{\partial \frac{\theta_L^{i*}}{x-\theta_L^{i*}}}{\partial e} = -A_1\frac{1+K(\kappa)}{1-\phi_2 q(e)}\big(\phi_1+\phi_2\big)\frac{\partial q}{\partial e} < 0; \quad \frac{\partial \frac{\theta_H^{i*}}{x-\theta_H^{i*}}}{\partial e} = -A_2\frac{1+K(\kappa)}{1-\phi_2 q(e)}\big(\phi_1+\phi_2\big)\frac{\partial q}{\partial e} < 0 \tag{1.16}$$

$$\frac{\partial \frac{\theta_L^{i*}}{x-\theta_L^{i*}}}{\partial \kappa} = -A_1\big(1+\phi_1 q(e)\big)\frac{\partial K(\kappa)}{\partial \kappa} < 0; \quad \frac{\partial \frac{\theta_H^{i*}}{x-\theta_H^{i*}}}{\partial \kappa} = -A_2\big(1+\phi_1 q(e)\big)\frac{\partial K(\kappa)}{\partial \kappa} < 0 \tag{1.17}$$

$$A_j = \frac{1}{\delta_j+1}\left[\frac{\psi_j}{1-\psi_j}\frac{1+\mathbb{K}(\kappa)}{1-\phi_2 q(e)}\frac{1+\phi_1 q(e)}{1+\tau}\right]^{-\frac{\delta_j+2}{\delta_j+1}}\frac{\psi_j}{1-\psi_j}\frac{1}{(1-\phi_2 q(e))(1+\tau)} > 0 \quad j\{1,2\}$$

1.3.3 Optimization problem: Government

The government maximizes the consumer's utility by choosing benefits κ and the level of enforcement e as a carrot and a stick. For those workers in formality there is a carrot which is benefits $K(\kappa)$, and there is a stick for those in informality which is the quality of

government enforcement ($q(e)$) and the level of fines ϕ_1.

The government equalizes the resources (from the social contribution in the formal jobs τ, the fines in the informal jobs $\phi_1 q(e)$ and a lump sum tax (T) with the spending *e.g.* the cost of the government enforcement and providing benefits which is given by the function $B(e, \kappa)$ (Equation 1.20)

Remark 2 *This maximization allows the government to choose the level of enforcement and benefits such that informality can exist in the labor market in its optimal choices. This possibility goes in the direction that some level of informality can be allowed by the government through the relaxation of the quality of the government enforcement and benefits. This point could be controversial if informality is perceived as a purely negative feature in the economy, although in this paper I use an agonistical approach to informality.*

$$\max_{e,\kappa} \quad U\big(C\big) \tag{1.18}$$

Subject to

$$C = \sum_s \left[\omega_s^i \theta_s^i \Big(1 - \phi_2 q(e)\Big) + \omega_s^f \theta_s^f \Big(1 + K(\kappa)\Big) \right] - T + \Pi \tag{1.19}$$

$$\sum_s \left[\tau \omega_s^f \theta_s^f + \phi_1 q(e) \theta_s^i \right] + T = B(e, \kappa) \quad s = \{L, H\} \tag{1.20}$$

F.O.C.

$$e: \quad \frac{\partial U}{\partial C} \left(\sum_s \omega_s^i \theta_s^i (\phi_1 - \phi_2) \frac{\partial q(e)}{\partial e} - \frac{\partial B(e, \kappa)}{\partial e} \right) = 0 \tag{1.21}$$

$$\kappa: \quad \frac{\partial U}{\partial C} \left(\sum_s \omega_s^f \theta_s^f \frac{\partial K(\kappa)}{\partial \kappa} - \frac{\partial B(e, \kappa)}{\partial \kappa} \right) = 0 \tag{1.22}$$

1.3.4 Equilibrium: Government

The functional forms which are chosen to estimate this model are such that the quality functions (enforcement e, and benefits κ) are increasing and concave where a and d are the quality parameters. The cost function of government enforcement and benefits are

quadratic

$$\begin{aligned} q(e) &= a\sqrt{e} \qquad K(\kappa) = d\sqrt{\kappa} \\ B(e,\kappa) &= b_1 e^2 + b_2 \kappa^2 + b_3 \end{aligned} \tag{1.23}$$

The equilibrium enforcement and benefits are:

$$e^* = \left(\frac{a(\phi_1 - \phi_2)\left(\omega_L^{i*}\theta_L^{i*} + \omega_H^{i*}\theta_H^{i*}\right)}{4b_1} \right)^{\frac{2}{3}}$$

The level of enforcement e depends positively on the mass of salaries in the informality $(\omega_L^i\theta_L^i + \omega_H^i\theta_H^i)$, the ratio of the quality and cost function parameters $\frac{a}{b_1}$, and the term $\phi_1 - \phi_2$, which is the difference between the fines which the firms pay for any informal job monitored and the worker loss in the informality. If the fines were equal to the workers' losses, the level of government enforcement would be zero because the effect of the informality is solved within the market, and the action of the government is not necessary.

The level of benefits κ that the government chooses depends positively on the mass of salaries in the formality $(\omega_L^f\theta_L^f + \omega_H^f\theta_H^f)$, and the ratio of the quality and cost function parameters $\frac{d}{b_2}$.

$$\kappa^* = \left(\frac{d(\omega_L^{f*}\theta_L^{f*} + \omega_H^{f*}\theta_H^{f*})}{4b_2} \right)^{\frac{2}{3}}$$

The level of taxes is given by the equalized budget constraint of the government:

$$T^* = \sum_s B(e^*, \kappa^*) - \tau\omega_s^{f*}\theta_s^{f*} - \phi_1\omega_s^{i*}\theta_s^{i*}q(e^*)$$

Definition 1 *Given the set of parameters, there is a unique equilibrium which determines the level of informality in each level of education (θ_s^{i*}) working, the level of government enforcement over informal jobs (e^*), the benefits that the government brings to the formal ones (κ^*), and the lump sum taxes collected from the households (T^*).*

In this section, I analyze what the effect is of the exogenous parameters in the main model equilibrium outcomes (θ_s^i, e and κ). I focus not only on the comparative statics

with respect to single parameters[11], but also considering the effect of a couple of them[12].

In the Equations in 1.24, I analyze the comparative statics of the equilibrium relative size of the informality (θ_s^{i*}) in respect to the fines (ϕ_1) and informal market imperfection or informal wage loss (ϕ_2). These shares depend negatively on the fines that the firm has to pay if the informal job is monitored (ϕ_1) and it is uncertain on the share of informal wage that the worker loses if their job is monitored (ϕ_2). Figure C.1 shows the simulated performance of a grid of ϕ_1 and ϕ_2, the informality is more sensitive with a change of ϕ_1 than ϕ_2.

Additionally, I consider the comparative statics of the informality shares respect to the quality function parameters. These shares decrease when the parameters of the quality of monitoring and benefits function (a and d) increase. The effect of a on the informality is higher than the effect of d, as is shown in Figure C.3. Figures C.4 and C.5, show the effect of the quality and cost parameters of the enforcement and the benefits, negative in quality and positive in costs.

$$\begin{array}{lll} \dfrac{\theta_s^{i*}}{\partial\phi_1} < 0 & \dfrac{\theta_s^{i*}}{\partial\phi_2} & \text{n.d.} \\ \dfrac{\theta_s^{i*}}{\partial a} < 0 & \dfrac{\theta_s^{i*}}{\partial d} < 0 & \end{array} \tag{1.24}$$

The comparative statics in the Equations in 1.25 show that government enforcement e^* is decreasing in the fines parameters (ϕ_1), decreasing in the quality of benefits (d) and unknown sign respect to the quality of government enforcement (a). The equilibrium behavior when these parameters change is also shown in Figures C.3 and C.4, in which the equilibrium is solved by fixing the other parameters. The effect of a on e^* is positive with this set of parameters, which was uncertain in the analytical analysis, and the effect of b_1 is negative (but the intensity also depend on the level of a as is shown in Figure C.4).

$$\frac{\partial e^*}{\partial\phi_1} < 0 \qquad \frac{\partial e^*}{\partial a} \quad \text{n.d.} \qquad \frac{\partial e^*}{\partial d} < 0 \tag{1.25}$$

[11] The analytical development is presented in Appendix C.

[12] The graphical analysis is presented in Appendix C, where there is a numerical exercise fixing the parameters as Argentina, and then computing the solution on a grid of two parameters. The parameters for Argentina are shown in Table 1.6.

$$\frac{\partial \kappa^*}{\partial \phi_1} > 0 \qquad \frac{\partial \kappa^*}{\partial a} > 0 \qquad \frac{\partial \kappa^*}{\partial d} > 0 \tag{1.26}$$

The benefits κ increases in fines (ϕ_1), the quality parameter of the enforcement (a) and the benefits (d), as is shown in Figure C.4. Figures C.3 and C.5, show the effect on κ of d and b_2, which are positive and negative respectively.

In the set of Equations in 1.27 and 1.28 I present the effect of the production function parameters on the model outcomes. In the case of the level of substitutability, the signs depend on the terms B_1 or B_2. If these terms are higher than one, a higher level of substitutability ($\delta_j \to -1$) leads to a lower level of informality, but if B_1 or B_2 are between 0 and 1, higher substitutability ($\delta_j \to -1$) leads to higher informality.

$$\begin{aligned}
&\frac{\theta_L^{i*}}{\partial \delta_2} > 0 \qquad \frac{e^*}{\partial \delta_2} > 0 \qquad \frac{\kappa^*}{\partial \delta_2} < 0 \\
&if \qquad B_2 = \left[\frac{\psi_2}{1-\psi_2} \frac{1+K(\kappa)}{1-\phi_2 q(e)} \frac{1+\phi_1 q(e)}{1+\tau} \right] > 1 \\
&\frac{\theta_L^{i*}}{\partial \delta_2} < 0 \qquad \frac{e^*}{\partial \delta_2} < 0 \qquad \frac{\kappa^*}{\partial \delta_2} > 0 \qquad if \quad 0 < B_2 < 1
\end{aligned} \tag{1.27}$$

$$\begin{aligned}
&\frac{\theta_H^{i*}}{\partial \delta_1} > 0 \qquad \frac{e^*}{\partial \delta_1} > 0 \qquad \frac{\kappa^*}{\partial \delta_1} < 0 \\
&if \qquad B_1 = \left[\frac{\psi_1}{1-\psi_1} \frac{1+K(\kappa)}{1-\phi_2 q(e)} \frac{1+\phi_1 q(e)}{1+\tau} \right] > 1 \\
&\frac{\theta_H^{i*}}{\partial \delta_1} < 0 \qquad \frac{e^*}{\partial \delta_1} < 0 \qquad \frac{\kappa^*}{\partial \delta_1} > 0 \qquad if \quad 0 < B_1 < 1
\end{aligned} \tag{1.28}$$

In the first two panels of Figure C.2, I observe how the share of informality changes with the level of substitutability between formal and informal workers in the case of the educated workers and the non educated ones (δ_1 and δ_2) and the shares in the production function (ψ_1 and ψ_2). If these shares were lower, B_j is lower than 1, then the sign of the comparative statics of the informality would be the opposite.

1.4 Results

The empirical strategy is to estimate both the production function and the quality and the cost functions. However, Botero García (2010) calibrates the elasticity of substitution

of a CES production function with informal and formal workers, the level of substitutability for low educated worker is -0.5, and between educated and non educated workers it is 3.3. Ulyssea (2010) also calibrates the parameters of the general level of substitutability between formal and informal production at -0.3. My empirical strategy is to estimate the parameters of the production function from the data. The production function is estimated by an approximation of a linear regression. The other parameters of the model are estimated with the Method of Moments.

In order to estimate the CES production function (and the nested production function) I use the linear Taylor-series approximation, which was first developed by Kmenta (1967).

$$y = \gamma\left[\psi_1 x_1^{-\alpha} + (1-\psi_1)x_2^{-\alpha}\right]^{\frac{-\upsilon}{\alpha}}$$

The second order Taylor Approximation at $\rho = 0$[13]

$$y = \gamma x_1^{\upsilon\psi_1} x_2^{\upsilon(1-\psi_1)} \exp(-0.5\alpha\upsilon\psi_1(1-\psi_1)(\ln x_1 - \ln x_2)^2)$$

In the case of my paper the production function is defined as follows, $y = \gamma l_H^{\rho_1} l_L^{\rho_2}$, $l_H = \left[\psi_1 (l_H^f)^{-\delta_1} + (1-\psi_1)(l_H^i)^{-\delta_1}\right]^{\frac{-(\upsilon_1)}{\delta_1}}$ and $l_L = \left[\psi_2 (l_L^f)^{-\delta_2} + (1-\psi_2)(l_L^i)^{-\delta_2}\right]^{\frac{-(\upsilon_2)}{\delta_2}}$, using the same methodology to estimate it we obtain that:

$$\begin{aligned}\ln y \approx \ln\gamma + \rho_1\upsilon_1\psi_1 \ln l_H^f + \rho_1\upsilon_1(1-\psi_1)\ln l_H^i - \frac{1}{2}\rho_1\upsilon_1\psi_1(1-\psi_1)\delta_1(\ln l_H^f - \ln l_H^i)^2 \\ + \rho_2\upsilon_2\psi_2 \ln l_L^f + \rho_2\upsilon_2(1-\psi_2)\ln l_L^i - \frac{1}{2}\rho_2\upsilon_2\psi_2(1-\psi_2)\delta_2(\ln l_L^f - \ln l_L^i)^2\end{aligned}$$

To estimate:

$$\ln y = \beta_0 + \beta_1 \ln l_H^f + \beta_2 \ln l_H^i + \beta_3(\ln l_H^f - \ln l_H^i)^2 + \beta_4 \ln l_L^f + \beta_5 \ln l_L^i + \beta_6(\ln l_L^f - \ln l_L^i)^2 + \epsilon$$

[13]Kmenta (1967) justifies this only by mathematical convenience and in order to estimate around the Cobb Douglass shape.

The main parameters of the production function are estimated:

$$\psi_1 = \frac{\beta_1}{\beta_1 + \beta_2}$$
$$\delta_1 = \frac{-2\beta_3}{\beta_2 \psi_1}$$
$$\psi_2 = \frac{\beta_4}{\beta_4 + \beta_5}$$
$$\delta_2 = \frac{-2\beta_6}{\beta_5 \psi_2}$$

	OLS	IV(1)	IV(2)
$\ln l_H^f$	2.608***	3.227***	3.687***
	(0.57)	(0.55)	(0.57)
$\ln l_H^i$	2.641***	3.493***	3.875***
	(0.47)	(0.42)	(0.51)
$\ln l_L^f$	2.031***	2.180***	1.649***
	(0.44)	(0.45)	(0.39)
$\ln l_L^i$	2.189***	2.312***	2.267***
	(0.46)	(0.48)	(0.42)
$(\ln l_H^f - \ln l_H^i)^2$	0.346***	0.472***	0.615***
	(0.08)	(0.07)	(0.09)
$(\ln l_L^f - \ln l_L^i)^2$	0.329***	0.375***	0.354***
	(0.08)	(0.08)	(0.07)
Constant	16.97***	17.67***	17.25***
	(0.79)	(0.71)	(0.77)
Year	yes	yes	yes
Sector	yes	yes	yes
R^2	0.573	0.689	0.667
N	987	807	865
ψ_1	0.497	0.48	0.488
δ_1	-0.528	-0.563	-0.651
ψ_2	0.481	0.485	0.421
δ_2	-0.624	-0.669	-0.741
$\rho_1 \upsilon_1$	5.25	6.72	7.56
$\rho_2 \upsilon_2$	4.22	4.49	3.92

Table 1.4: Estimation of the production function

The results of the production function estimation appear in Table 1.4, the dependent variable is the logarithm of the aggregate value by sector[14] and the independent variables

[14]The seven sector that I consider are: primary, manufacturing industries, construction, commerce,

are the logarithm of the informality shares. In Table 1.4, I estimate the parameters for all the countries together. The level of substitutability ($\delta's$) is slightly higher in the case of the non educated workers, but even for the educated it is relatively high. The shares of workers are lower than 0.5, so the terms B_1 and B_2 are closer to 1, if these terms are lower than 1, the informality is decreasing with respect to δ_1 and δ_2.

The second and third columns present the IV estimation instrumented by the lags. In the second column, the estimation is through the inclusion of the error term of the first steps. Note that the joint test of these error terms is significative[15]. In the third column, the estimation is instrumented by the predicted estimation from the first step. In the next section, I will use the estimation from the second column.

In the estimation of the model's parameters through the Methods of Moments, minimizing the distance between the moments from the model and the data, I match the quality of government enforcement $q(e)$, the quality of the benefits that the formal workers receive from the government $K(\kappa)$, the level of informality in both levels of education, the level of lump sum taxes, and the share of public expenditure.

As is discussed in the introduction there are no clear variables to match in the case of the quality functions, so I want to compare how the fit is in the different countries[16]. In the case of the quality of government enforcement I match the indicator with *compliance with the law*, and the quality of the benefits is matched with the indicator *taxes well spent*. Informality shares are estimated with the National Household Surveys, the public expenditures are matched with the figure from the National Accounts, and the level of the lump sum tax is matched at zero[17].

Firstly, I present the result of the estimation for each country independently. The model fits the data quite well in general and in particular the ranking of the countries is respected, as is presented in Table 1.5. This latter point is somewhat important given the comparative objective of this paper, principally, if the comparison is between the countries with better performance, such as Brazil and Uruguay, with those with low performance as in the case of Peru. However, there are some features that I want to highlight: informality

transport, finance and services.

[15] F(4, 775) = 9.41 $Prob > F = 0.00$.

[16] The variables to fit are taken by surveys which are published for the the IADB.

[17] The wages are not considered as moments because there is no data about work hours in Colombia, and there is a lack of information to construct (comparatively) the formal wages including benefits for all the countries.

for the low educated workers is underestimated, and probably the main problem is the overestimation in the high educated ones. The quality of government enforcement and quality of benefits are well estimated with the exception of the government enforcement in Brazil. The share of public expenditure is well estimated, and the taxes, which is a residual variable, fit well at zero.

	Argentina		Brasil		Colombia		Peru		Uruguay	
	Model	Data	Model	Data	Model	Data	Model	Data	Model	Data
θ_L^{i*}	0.2616	0.338	0.2193	0.228	0.2343	0.349	0.4134	0.485	0.2266	0.240
θ_H^{i*}	0.1244	0.079	0.0668	0.026	0.1151	0.057	0.1740	0.117	0.0849	0.024
$q(e^*)$	0.2002	0.232	0.2907	0.431	0.3165	0.317	0.1574	0.162	0.5080	0.508
$K(\kappa^*)$	0.1884	0.209	0.1458	0.161	0.1602	0.167	0.0922	0.101	0.3575	0.358
T	0.0001	0.000	-0.0003	0.000	-0.0002	0.000	0.0003	0.000	0.0001	0.000
SCP	0.1770	0.162	0.2018	0.246	0.1708	0.213	0.1237	0.110	0.1505	0.151

Table 1.5: Moment matching with the country by country estimation.

In Table 1.6 the estimation of the parameters for all five countries can be observed. Given that the effect of the quality parameters and the enforcement go in the same direction, it is not possible to observe a clear ranking in them. However, comparing a group of parameters as the term $\left(\frac{a(\phi_1-\phi_2)}{4b_1}\right)$ (which multiply government enforcement e) is higher in countries with a lower level of informality, especially when the extreme countries (Uruguay or Brazil and Peru) are compared. Considering the countries with a low level of informality, the parameters of Uruguay and Brazil are 6.5 and 5.3 respectively and the Peruvian is 3.3. The order when the parameters that multiply the level of benefits are considered $(\frac{d}{4b_2})$ is not so clear. Again, the parameters of Peru is the lowest one (1.08), while the parameters are quite similar for Argentina, Colombia and Uruguay (2.45, 2.14 and 2.22 respectively).

In order to disentangle how each parameter can differentiate the optimal level of the quality functions and the level of the informality for these five countries, I will estimate the parameters for all countries together, leaving only one parameter being different country by country. Tables A.1 - A.5 show the moment matching when the fines ϕ_1, the market imperfection ϕ_2, quality of the government enforcement a, the quality of the benefits d, and the cost parameters b_1 and b_2 are different country by country using the levels of informality, the quality of enforcement and benefits, the lump sum tax and public expenditure as moments. Obviously, the model fits worse than when the countries all have different parameters.

	Argentina	Brazil	Colombia	Peru	Uruguay
a	0.8951	0.9180	1.5336	0.6329	1.9675
	(*0.000*)	(*0.096*)	(*0.000*)	(*0.002*)	(*0.002*)
d	0.8426	0.6521	0.8087	0.6247	1.6775
	(*0.000*)	(*0.001*)	(*0.001*)	(*0.037*)	(*0.007*)
b_1	0.0944	0.0645	0.1138	0.0001	0.0209
	(*0.186*)	(*0.050*)	(*0.001*)	(*0.002*)	(*0.001*)
b_2	0.0861	0.1159	0.0943	0.1442	0.1888
	(*0.053*)	(*0.311*)	(*0.000*)	(*0.027*)	(*0.001*)
b_3	0.001	0.001	0.001	0.001	0.001
	(−)	(−)	(−)	(−)	(−)
ϕ_1	1.1710	1.6423	1.2326	0.1319	0.3776
	(*0.327*)	(*0.131*)	(*0.065*)	(*0.005*)	(*0.142*)
ϕ_2	0.1878	0.1547	0.0000	0.1298	0.1016
	(*0.192*)	(*0.004*)	(*0.000*)	(*0.127*)	(*0.002*)

Table 1.6: Estimation of all quality and cost parameters with the production function's parameters $\delta_1 = -0.563$, $\delta_2 = -0.669$, $\psi_1 = 0.480$, $\psi_2 = 0.485$, $\rho_1 \upsilon_1 = 6.72$ and $\rho_2 \upsilon_2 = 4.92$. The standard deviations are estimated by the gradient of the moments vector.

Table A.1 shows the model matching when the fines ϕ_1 are different country by country, the high informality is the moment which has most problems to fit with a clear overestimation (about double of the data). Regarding the fit with the other moments, there is no clear tendency (neither underestimation nor overestimation for all countries).

Table A.2 shows the moment matching when the market imperfection ϕ_2 is different country by country, the high informality is again overestimated, and the low informality is well estimated for those countries with low levels (Brazil and Uruguay) but there is an underestimation for the other ones.

In the left panel of Table 1.7, I observe the performance of ϕ_1, which is the amount of fines that the firm would have to pay to have workers in informality, leaving the other parameters constant country by country. There is a clear and direct relation between the level of fines and level of informality, in particular it is higher in Brazil and Uruguay, 1.4 and 0.88, than in Argentina and Peru, 0.2 and 0.025 respectively. In the case of Colombia, it has a relatively high level of fines that is reflected in lower levels of informality in low educated workers (see Table A.1). The ability of ϕ_1 to capture heterogeneity in the model is a feature which is is line with the empirical evidence of the literature.

Parameters	Value	Std. Dev.	Parameters	Value	Std. Dev.
a	1.1045	(0.031)	a	0.2581	(0.008)
d	0.6185	(0.123)	d	0.3625	(0.124)
b_1	0.0109	(0.0001)	b_1	0.0000	(0.000)
b_2	0.0297	(0.005)	b_2	0.0070	(0.001)
b_3	0.001	$(-)$	b_3	0.001	$(-)$
ϕ_1^a	0.2012	(0.008)	ϕ_1	0.8716	(0.037)
ϕ_1^b	1.3991	(0.027)	ϕ_2^a	0.8528	(0.041)
ϕ_1^c	0.9332	(0.017)	ϕ_2^b	0.1104	(0.293)
ϕ_1^p	0.0250	(0.001)	ϕ_2^c	0.0122	(0.003)
ϕ_1^u	0.8836	(0.020)	ϕ_2^p	0.8546	(0.035)
ϕ_2	0.0000	(0.0004)	ϕ_2^u	0.1421	(0.297)

Table 1.7: Estimation of fines (ϕ_1) and market imperfection (ϕ_2) parameters with the production function's parameters $\delta_1 = -0.563$, $\delta_2 = -0.669$, $\psi_1 = 0.480$, $\psi_2 = 0.485$, $\rho_1 \upsilon_1 = 6.72$ and $\rho_2 \upsilon_2 = 4.92$. The standard deviations are estimated by the gradient of the moments vector.

In the right panel of Table 1.7, there is the parameter estimation when ϕ_2 (market imperfection) is different country by country. Countries with high levels of informality, such as Argentina and Peru present high values of ϕ_2 and Brazil and Uruguay have lower values. In this case, Colombia is again the country in which the order does not fit, and it is reflected in the underestimation in the informality of low educated workers (see Table A.2). The market imperfection parameter is not able to capture the heterogeneity as the fines parameter was able to.

In Table 1.8, I present the estimation when the parameters of the quality functions change. The moment matching of these estimation are presented in Tables A.3 and A.4. The matching which corresponds to the a estimation (Table A.3) shows the overestimation of the high educated informality and the underestimation of the low educated informality for Brazil and Uruguay, but a good fit for the other countries. The matching of d estimation (Table A.4) is similar to the previous one. Both quality parameters allow the model to capture the heterogeneity in the moments.

The left panel of Table 1.8 shows the estimation of the quality of government enforcement, and the right panel the quality of the benefits. The estimations of the quality parameters show the same order as the informality and not the order of the quality moments among countries. These estimations show that better quality parameters means

Parameters	Value	Std. Dev.	Parameters	Value	Std. Dev.
a^a	0.3226	(0.030)	a	0.3280	(0.088)
a^b	0.7757	(0.242)	d^a	0.2866	(0.003)
a^c	0.2688	(0.010)	d^b	0.3640	(0.004)
a^p	0.0581	(0.093)	d^c	0.1418	(0.001)
a^u	0.8792	(0.039)	d^p	0.0220	(0.0003)
d	0.6679	(0.019)	d^u	0.5059	(0.01)
b_1	0.0012	(0.00002)	b_1	0.0002	(0.00001)
b_2	0.0449	$(0.0002$	b_2	0.0007	(0.00001)
b_3	0.001	$(-)$	b_3	0.001	$(-)$
ϕ_1	1.132	(0.065)	ϕ_1	0.6368	(0.015)
ϕ_2	0.0152	(0.007)	ϕ_2	0.0898	(0.0061)

Table 1.8: Estimation of quality function parameters with the production function's parameters $\delta_1 = -0.563$, $\delta_2 = -0.669$, $\psi_1 = 0.480$, $\psi_2 = 0.485$, $\rho_1 \upsilon_1 = 6.72$ and $\rho_2 \upsilon_2 = 4.92$. The standard deviations are estimated by the gradient of the moments vector.

lower informality.

Table 1.9 shows the estimation of the cost function, and Table A.5 is the moment matching of this estimation. The cost function is able to reproduce the heterogeneity, the model matches the order well but it has some problems in the case of Argentina. Lower cost function parameters lead to low levels of informality.

1.5 Concluding Remarks

I consider informality using a loose form of dualism and it is formalized through the definition of the production function. This is one of the main contributions of this paper.

The model captures the quality of government enforcement and benefits, and the informality for two levels of education in five different Latin American countries which have a great heterogeneity among them.

The first result is the estimation of a production function where formal and informal workers coexist when there are two levels of education. I estimate the production function using data from all countries, and the level of substitutability of the low educated workers is higher than the level of the high educated workers.

The second result is the estimation of the parameters in the quality of government

Parameters	Value	Std Dev
a	0.8707	(0.019)
d	0.8911	(0.049)
b_1^a	0.0023	(0.002)
b_2^a	0.1286	(0.017)
b_1^b	0.0019	(0.001)
b_2^b	0.0192	(0.000)
b_1^c	0.3458	(0.080)
b_2^c	0.4641	(0.212)
b_1^p	1.7367	(0.167)
b_2^p	0.1863	(0.028)
b_1^u	0.0000	(0.000)
b_2^u	0.0087	(0.000)
b_3	0.001	$(-)$
ϕ_1	1.2232	(0.192)
ϕ_2	0.0002	(0.006)

Table 1.9: Estimation of cost function parameters with the production function's parameters $\delta_1 = -0.563$, $\delta_2 = -0.669$, $\psi_1 = 0.480$, $\psi_2 = 0.485$, $\rho_1 \upsilon_1 = 6.72$ and $\rho_2 \upsilon_2 = 4.92$. The standard deviations are estimated by the gradient of the moments vector.

enforcement, the quality of the benefits that the workers receive in formal employment, the fines, the market imperfection and the cost function by the Method of Moments. Through this estimation, the model can capture the ranking of countries as well as the informality for different educative levels. However, the model has some difficulties in capturing the level of informality for high educative levels.

The third result is the estimation leaving one parameter free, only the market imperfection parameter ϕ_2 does not allow the model to reproduce the heterogeneity. The fines (ϕ_1), quality parameters (a and d) and cost parameters (b_1 and b_2) allow the model to generate heterogeneity, and the moments are matched quite well.

In short, the model captures the main features of these economies and produces a good estimation of the parameters describing the countries' heterogeneity.

1.6 Appendix: Tables and figures.

The Tables A.1 - A.4 show the model matching when the estimation is made leaving only one parameter (ϕ_1, ϕ_2, a, d, b_1 and b_2) different among countries.

Table A.1: Moment matching leaving free the fine parameter ϕ_1.

	Argentina		Brasil		Colombia		Peru		Uruguay	
	Model	Data	Model	Data	Model	Data	Model	Data	Model	Data
θ_L^{i*}	0.3731	0.338	0.1817	0.228	0.2363	0.349	0.4040	0.485	0.2436	0.240
θ_H^{i*}	0.1609	0.079	0.0964	0.026	0.1158	0.057	0.1710	0.117	0.1182	0.024
$q(e^*)$	0.3075	0.232	0.4041	0.431	0.4069	0.317	0.1589	0.162	0.4061	0.508
$K(\kappa^*)$	0.1582	0.209	0.1613	0.161	0.1639	0.167	0.1546	0.101	0.1641	0.358
T	0.0003	0.000	-0.0002	0.000	-0.0002	0.000	0.0005	0.000	-0.0002	0.000
SCP	0.1338	0.162	0.1948	0.246	0.1718	0.213	0.1267	0.110	0.1692	0.151

Table A.2: Moment matching leaving the market imperfection ϕ_2 free .

	Argentina		Brasil		Colombia		Peru		Uruguay	
	Model	Data	Model	Data	Model	Data	Model	Data	Model	Data
θ_L^{i*}	0.2784	0.338	0.2214	0.228	0.2599	0.349	0.2822	0.485	0.2176	0.240
θ_H^{i*}	0.1300	0.079	0.1106	0.026	0.1238	0.057	0.1312	0.117	0.1092	0.024
$q(e^*)$	0.1532	0.232	0.4521	0.431	0.3964	0.317	0.1488	0.162	0.4436	0.508
$K(\kappa^*)$	0.1305	0.209	0.1302	0.161	0.1310	0.167	0.1305	0.101	0.1301	0.358
T	-0.0000	0.000	-0.0002	0.000	-0.0004	0.000	-0.0000	0.000	-0.0002	0.000
SCP	0.1419	0.162	0.1737	0.246	0.1505	0.213	0.1410	0.110	0.1741	0.151

Table A.3: Moment matching leaving the quality of the government enforcement parameter a free.

	Argentina		Brasil		Colombia		Peru		Uruguay	
	Model	Data	Model	Data	Model	Data	Model	Data	Model	Data
θ_L^{i*}	0.2983	0.338	0.1842	0.228	0.3173	0.349	0.3946	0.485	0.1674	0.240
θ_H^{i*}	0.1366	0.079	0.0973	0.026	0.1428	0.057	0.1679	0.117	0.0910	0.024
$q(e^*)$	0.1952	0.232	0.4871	0.431	0.1578	0.317	0.0227	0.162	0.5467	0.508
$K(\kappa^*)$	0.1578	0.209	0.1559	0.161	0.1570	0.167	0.1504	0.101	0.1545	0.358
T	-0.0001	0.000	-0.0002	0.000	-0.0000	0.000	0.0004	0.000	-0.0001	0.000
SCP	0.1524	0.162	0.1931	0.246	0.1474	0.213	0.1291	0.110	0.2020	0.151

Table A.4: Moment matching leaving the quality of the benefits d free.

	Argentina		Brasil		Colombia		Peru		Uruguay	
	Model	Data	Model	Data	Model	Data	Model	Data	Model	Data
θ_L^{i*}	0.2759	0.338	0.2481	0.228	0.3284	0.349	0.3630	0.485	0.2021	0.240
θ_H^{i*}	0.1291	0.079	0.1198	0.026	0.1464	0.057	0.1577	0.117	0.1038	0.024
$q(e^*)$	0.3032	0.232	0.2926	0.431	0.3195	0.317	0.3278	0.162	0.2721	0.508
$K(\kappa^*)$	0.2079	0.209	0.2861	0.161	0.0803	0.167	0.0066	0.101	0.4398	0.358
T	-0.0000	0.000	0.0001	0.000	-0.0002	0.000	-0.0003	0.000	0.0004	0.000
SCP	0.1544	0.162	0.1621	0.246	0.1444	0.213	0.1408	0.110	0.1789	0.151

Table A.5: Moment matching leaving the cost function b_1 and b_2 free.

	Argentina		Brasil		Colombia		Peru		Uruguay	
	Model	Data	Model	Data	Model	Data	Model	Data	Model	Data
θ_L^{i*}	0.1795	0.338	0.1420	0.228	0.3553	0.349	0.3670	0.485	0.1140	0.240
θ_H^{i*}	0.0956	0.079	0.0812	0.026	0.1552	0.057	0.1590	0.117	0.0699	0.024
$q(e^*)$	0.4694	0.232	0.4572	0.431	0.1243	0.317	0.0745	0.162	0.4933	0.508
$K(\kappa^*)$	0.1609	0.209	0.2961	0.161	0.1041	0.167	0.1402	0.101	0.3766	0.358
T	-0.0002	0.000	0.0001	0.000	-0.0000	0.000	0.0002	0.000	0.0001	0.000
SCP	0.1959	0.162	0.2162	0.246	0.1411	0.213	0.1363	0.110	0.2152	0.151

1.7 Appendix: Household problem

The complete F.O.C. of the Representative household problem are:

$$C\frac{\partial L}{\partial C} = 0: \quad C\big(\frac{\partial U\big(C\big)}{\partial C} + \lambda_1^{HH}\big) = 0 \tag{B.1}$$

$$\theta_L^i \frac{\partial L}{\partial \theta_L^i} = 0: \quad \theta_L^i\bigg(- \lambda_1^{HH}\bigg(\omega_L^i\Big(1 - \phi_2 q(e)\Big)\bigg) + \lambda_2^{HH}\bigg) = 0 \tag{B.2}$$

$$\theta_L^f \frac{\partial L}{\partial \theta^f} = 0: \quad \theta_L^f\bigg(- \lambda_1^{HH}\bigg(\omega_L^f\big(1 + K(\kappa)\big)\bigg) + \lambda_2^{HH}\bigg) = 0 \tag{B.3}$$

$$\theta_H^i \frac{\partial L}{\partial \theta_H^i} = 0: \quad \theta_H^i\bigg(- \lambda_1^{HH}\bigg(\omega_H^i\Big(1 - \phi_2 q(e)\Big)\bigg) + \lambda_3^{HH}\bigg) = 0 \tag{B.4}$$

$$\theta_H^f \frac{\partial L}{\partial \theta^f} = 0: \quad \theta_H^f\bigg(- \lambda_1^{HH}\bigg(\omega_H^f\big(1 + K(\kappa)\big)\bigg) + \lambda_3^{HH}\bigg) = 0 \tag{B.5}$$

$$\lambda_1^{HH}\frac{\partial L}{\partial \lambda_1^{HH}} = 0: \quad \lambda_1^{HH}\bigg(C - \sum_s \bigg[\omega_s^i\theta_s^i\Big(1 - \phi_2 q(e)\Big) - \omega_s^f\theta_s^f\Big(1 + K(\kappa)\Big)\bigg] + T - \Pi\bigg) = 0 \tag{B.6}$$

$$\lambda_2^{HH}\frac{\partial L}{\partial \lambda_2^{HH}} = 0: \quad \lambda_2^{HH}\Big(- x + \theta_L^i + \theta_L^f\Big) = 0 \tag{B.7}$$

$$\lambda_3^{HH}\frac{\partial L}{\partial \lambda_3^{HH}} = 0: \quad \lambda_3^{HH}\Big(x - 1 + \theta_H^i + \theta_H^f\Big) = 0 \tag{B.8}$$

1.8 Appendix: Comparative Statics of the equilibrium variables

In this section I show the comparative statics of the informality shares, the level of government enforcement and the benefits (θ_L^i, θ_H^i, e and κ) and the main exogenous variables (ϕ_1, ϕ_2, δ_j, ψ_j, a and d).

Some terms for $j = \{1, 2\}$:

$$A_j = \frac{1}{\delta_j + 1}\left[\frac{\psi_j}{1-\psi_j}\frac{1+\mathbb{K}(\kappa)}{1-\phi_2 q(e)}\frac{1+\phi_1 q(e)}{1+\tau}\right]^{-\frac{\delta_j+2}{\delta_j+1}}\frac{\psi_j}{1-\psi_j}\frac{1}{(1-\phi_2 q(e))(1+\tau)} > 0$$

$$E_1 = \left(\omega_L^i\theta_L^i + \omega_H^i\theta_H^i\right)\left(\phi_1 - \phi_2\right)\frac{\partial^2 q}{\partial e^2} - \frac{\partial^2 B}{\partial e^2} < 0$$

$$K_1 = \left(\omega_L^f(x-\theta_L^i) + \omega_H^f(1-x-\theta_H^i)\right)\frac{\partial^2 K}{\partial \kappa^2} - \frac{\partial^2 B}{\partial \kappa^2} < 0$$

$$B_j = \left[\frac{\psi_j}{1-\psi_j}\frac{1+K(\kappa)}{1-\phi_2 q(e)}\frac{1+\phi_1 q(e)}{1+\tau}\right] > 0$$

$$C_j = \frac{1}{\delta_j + 1}\left[\frac{\psi_j}{1-\psi_j}\frac{1+K(\kappa)}{1-\phi_2 q(e)}\frac{1+\phi_1 q(e)}{1+\tau}\right]^{-\frac{\delta_j+2}{\delta_j+1}}\frac{1+K(\kappa)}{(1-\phi_2 q(e))}\frac{1+\phi_1 q(e)}{(1+\tau)} > 0$$

$$j = \{1, 2\}$$

Denominator D:

$$D = \overbrace{\frac{x}{\left(x-\theta_L^i\right)^2}\left[\frac{1-x}{\left(1-x-\theta_H^i\right)^2}\right]E_1 K_1}^{\geq 0} + \left[\frac{A_1 x}{\left(x-\theta_L^i\right)^2} + \frac{A_2(1-x)}{\left(1-x-\theta_H^i\right)^2}\right] * \left[\underbrace{\left(\omega_L^f + \omega_H^f\right)\left(1+\phi_1 q(e)\right)\left[\frac{\partial K}{\partial \kappa}\right]^2 E_1}_{\leq 0} \underbrace{-\left(\omega_L^i + \omega_H^i\right)\left(\frac{1+K(\kappa)}{1-\phi_2 q(e)}\right)\left(\phi_1^2 - \phi_2^2\right)\left[\frac{\partial q}{\partial e}\right]^2 K_1}_{\geq 0}\right] \tag{C.1}$$

In the interval of interest of the parameters the simulation shows that the positive terms

are higher than the negative one, then:

$$D > 0$$

1.8.1 Comparative statics of θ_L^{i*} and θ_H^{i*}

In this section I show the comparative statics of θ_L^i and θ_H^i with respect to the main exogenous parameters of the model. The change of θ_L^i and θ_H^i with respect to ϕ_1 is negative, the amount of fines impact negatively on the equilibrium informal shares:

Numerator N_{11}:

$$\frac{\partial \theta_L^i}{\partial \phi_1} = \frac{N_{11}}{D} < 0 \tag{C.2}$$

$$N_{11} = A_2\Big(1 + K(\kappa)\Big)\frac{1-x}{(1-x-\theta_H^i)^2}K_1\left[\frac{(\phi_1+\phi_2)}{1-\phi_2 q(e)}\big(\omega_L^i\theta_L^i + \omega_H^i\theta_H^i\big)\Big[\frac{\partial q}{\partial e}\Big]^2 - q(e)E_1\right] < 0 \tag{C.3}$$

Numerator N_{21}:

$$\frac{\partial \theta_H^i}{\partial \phi_1} = \frac{N_{21}}{D} < 0 \tag{C.4}$$

$$N_{21} = A_1\Big(1 + K(\kappa)\Big)\frac{x}{(x-\theta_L^i)^2}K_1\left[\frac{(\phi_1+\phi_2)}{1-\phi_2 q(e)}\big(\omega_L^i\theta_L^i + \omega_H^i\theta_H^i\big)\Big[\frac{\partial q}{\partial e}\Big]^2 - q(e)E_1\right] < 0 \tag{C.5}$$

The impact is higher in the lower (higher) skilled informal worker if:

$$\frac{\partial \theta_L^i}{\partial \phi_1} \geq \frac{\partial \theta_H^i}{\partial \phi_1} \qquad \text{iff} \quad \frac{A_2(1-x)}{\big(1-x-\theta_H^i\big)^2} \geq \frac{A_1 x}{\big(x-\theta_L^i\big)^2}$$

In the case of the comparative statics of θ_L^i and θ_H^i respect to ϕ_2 the sign is uncertain.

Numerator N_{12}:

$$\frac{\partial \theta_L^i}{\partial \phi_2} = \frac{N_{12}}{D} \quad \text{n.d.} \tag{C.6}$$

$$N_{12} = -A_2\frac{1+K(\kappa)}{1-\phi_2 q(e)}\frac{1-x}{(1-x-\theta_H^i)^2}K_1\left[\big(1+\phi_1 q(e)\big)q(e)E_1 + (\phi_1+\phi_2)\big(\omega_L^i\theta_L^i + \omega_H^i\theta_H^i\big)\Big[\frac{\partial q}{\partial e}\Big]^2\right] \quad \text{n.d.} \tag{C.7}$$

Numerator N_{22}:

$$\frac{\partial \theta_H^i}{\partial \phi_2} = \frac{N_{22}}{D} \quad \text{n.d.} \tag{C.8}$$

$$N_{22} = -A_1 \frac{1+K(\kappa)}{1-\phi_2 q(e)} \frac{x}{(x-\theta_L^i)^2} K_1 \left[\Big(1+\phi_1 q(e)\Big) q(e) E_1 + \big(\phi_1+\phi_2\big)\big(\omega_L^i \theta_L^i + \omega_H^i \theta_H^i\big) \Big[\frac{\partial q}{\partial e}\Big]^2 \right] \quad \text{n.d.} \tag{C.9}$$

Analyzing the comparative statics of θ_L^i with respect to δ_2, the sign depends on B_1. If B_1 is higher than 1, the effect of a higher level of substitutability impacts positively on the informality. in the case of θ_H^i the sign is uncertain.

Numerator N_{13}:

$$\begin{aligned} \frac{\partial \theta_L^i}{\partial \delta_2} &= \frac{N_{13}}{D} > 0 \quad if \quad B_2 > 1 \\ \frac{\partial \theta_L^i}{\partial \delta_2} &= \frac{N_{13}}{D} < 0 \quad if \quad B_2 < 1 \end{aligned} \tag{C.10}$$

$$N_{13} = \frac{\ln B_2 B_2^{\frac{-1}{\delta_2+1}}}{(\delta_2+1)^2} * \underbrace{\left[\frac{1-x}{(1-x-\theta_H^i)^2} E_1 K_1 - A_1 \frac{1+K(\kappa)}{1-\phi_2 q(e)} \big(\phi_1^2 - \phi_2^2\big) \Big[\frac{\partial q}{\partial e}\Big]^2 \omega_H^i K_1 + A_1 \big(1+\phi_1 q(e)\big) \Big[\frac{\partial K}{\partial \kappa}\Big]^2 \omega_H^f E_1 \right]}_{>0 \quad \text{(by simulation)}}$$

$$if \quad B_2 > 1 \tag{C.11}$$

Numerator N_{23}:

$$\frac{\partial \theta_H^i}{\partial \delta_2} = \frac{N_{23}}{D} \quad \text{n.d.} \tag{C.12}$$

$$N_{23} = A_1 \frac{\ln B_2 B_2^{\frac{-1}{\delta_2+1}}}{(\delta_2+1)^2} \left[\frac{1+K(\kappa)}{1-\phi_2 q(e)} \big(\phi_1^2 - \phi_2^2\big) \Big[\frac{\partial q}{\partial e}\Big]^2 \omega_L^i K_1 - \big(1+\phi_1 q(e)\big) \omega_L^f \Big[\frac{\partial K}{\partial \kappa}\Big]^2 E_1 \right] \quad \text{n.d.} \tag{C.13}$$

When ψ_2 changes, impacts negatively on the low skilled informality and has an uncertain effect on the higher ones:

Numerator N_{14}:

$$\frac{\partial \theta_L^i}{\partial \psi_2} = \frac{N_{14}}{D} < 0 \tag{C.14}$$

$$N_{14} = -\frac{C_2}{(1-\psi_2)^2} * \left[\frac{1-x}{(1-x-\theta_H^i)^2} E_1 K_1 - A_1 \frac{1+K(\kappa)}{1-\phi_2 q(e)} \big(\phi_1^2 - \phi_2^2\big) \Big[\frac{\partial q}{\partial e}\Big]^2 \omega_H^i K_1 + A_1 \big(1+\phi_1 q(e)\big) \Big[\frac{\partial K}{\partial \kappa}\Big]^2 \omega_H^f E_1 \right] < 0 \tag{C.15}$$

Numerator N_{24}:

$$\frac{\partial \theta_H^i}{\partial \psi_2} = \frac{N_{24}}{D} \quad \text{n.d.} \tag{C.16}$$

$$N_{24} = -\frac{C_2}{(1-\psi_2)^2} \left[\frac{1+K(\kappa)}{1-\phi_2 q(e)} \big(\phi_1^2 - \phi_2^2\big) \Big[\frac{\partial q}{\partial e}\Big]^2 \omega_L^i K_1 - \big(1+\phi_1 q(e)\big) \omega_L^f \Big[\frac{\partial K}{\partial \kappa}\Big]^2 E_1 \right] \quad \text{n.d.} \tag{C.17}$$

The sign of θ_H^i when δ_2, changes depends on B_2. If B_2 is higher than 1, the effect of a higher level of substitutability impacts positively on the informality. in the case of θ_L^i the sign is uncertain.

Numerator N_{15}:

$$\frac{\partial \theta_L^i}{\partial \delta_1} = \frac{N_{15}}{D} \quad \text{n.d.} \tag{C.18}$$

$$N_{15} = \frac{\ln B_1 B_1^{\frac{-1}{\delta_1+1}}}{(\delta_1+1)^2} \left[A_2 \frac{1+K(\kappa)}{1-\phi_2 q(e)} (\phi_1^2 - \phi_2^2) \Big[\frac{\partial q}{\partial e}\Big]^2 \omega_L^i K_1 - A_2 \big(1+\phi_1 q(e)\big) \omega_H^f \Big[\frac{\partial K}{\partial \kappa}\Big]^2 E_1 \right] \quad \text{n.d.} \tag{C.19}$$

Numerator N_{25}:

$$\begin{aligned} \frac{\partial \theta_H^i}{\partial \delta_1} &= \frac{N_{25}}{D} > 0 \quad if \quad B_1 > 1 \\ \frac{\partial \theta_H^i}{\partial \delta_1} &= \frac{N_{25}}{D} < 0 \quad if \quad B_1 < 1 \end{aligned} \tag{C.20}$$

$$N_{25} = \frac{\ln B_1 B_1^{\frac{-1}{\delta_1+1}}}{(\delta_1+1)^2} * \left[\frac{1-x}{(1-x-\theta_H^i)^2} E_1 K_1 - A_2 \frac{1+K(\kappa)}{1-\phi_2 q(e)} \big(\phi_1^2 - \phi_2^2\big) \Big[\frac{\partial q}{\partial e}\Big]^2 \omega_L^i K_1 + A_2 \big(1+\phi_1 q(e)\big) \Big[\frac{\partial K}{\partial \kappa}\Big]^2 \omega_L^f E_1 \right]$$

$$N_{25} > 0$$

$$if \quad B_1 > 1 \tag{C.21}$$

The change of ψ_1 has an uncertain effect on θ_L^i, and a negative effect on θ_H^i:

Numerator N_{16}:

$$\frac{\partial \theta_L^i}{\partial \psi_1} = \frac{N_{16}}{D} \quad \text{n.d.} \tag{C.22}$$

$$N_{16} = -\frac{C_1}{(1-\psi_1)^2} \left[A_2 \frac{1+K(\kappa)}{1-\phi_2 q(e)} (\phi_1^2 - \phi_2^2) \Big[\frac{\partial q}{\partial e}\Big]^2 \omega_L^i K_1 - A_2 \big(1+\phi_1 q(e)\big) \omega_H^f \Big[\frac{\partial K}{\partial \kappa}\Big]^2 E_1 \right] \quad \text{n.d.} \tag{C.23}$$

Numerator N_{26}:

$$\frac{\partial \theta_H^i}{\partial \psi_1} = \frac{N_{26}}{D} < 0 \tag{C.24}$$

$$N_{26} = -\frac{C_1}{(1-\psi_1)^2} * \left[\frac{1-x}{(1-x-\theta_H^i)^2} E_1 K_1 - A_2 \frac{1+K(\kappa)}{1-\phi_2 q(e)} \big(\phi_1^2 - \phi_2^2\big) \Big[\frac{\partial q}{\partial e}\Big]^2 \omega_L^i K_1 + A_2 \big(1+\phi_1 q(e)\big) \Big[\frac{\partial K}{\partial \kappa}\Big]^2 \omega_L^f E_1 \right] < 0 \tag{C.25}$$

The change of the quality parameters a and d on θ_L^i and θ_H^i is negative. If the quality parameters are higher the informal shares are reduced:

Numerator N_{17}:

$$\frac{\partial \theta_L^i}{\partial a} = \frac{N_{17}}{D} < 0 \tag{C.26}$$

$$N_{17} = A_2 \frac{1+K(\kappa)}{1-\phi_2 q(e)} \frac{1-x}{(1-x-\theta_H^i)^2} \Big(\phi_1+\phi_2\Big) \left[-\frac{\partial q}{\partial a} + \big(\omega_L^i \theta_L^i + \omega_H^i \theta_H^i\big)\big(\phi_1-\phi_2\big) \frac{\partial^2 q}{\partial e \partial a} K_1 \right] < 0 \tag{C.27}$$

Numerator N_{27}:

$$\frac{\partial \theta_H^i}{\partial a} = \frac{N_{27}}{D} < 0 \tag{C.28}$$

$$N_{27} = A_1 \frac{1+K(\kappa)}{1-\phi_2 q(e)} \frac{x}{(x-\theta_L^i)^2} \Big(\phi_1+\phi_2\Big) K_1 \left[\big(\omega_L^i \theta_L^i + \omega_H^i \theta_H^i\big)\big(\phi_1-\phi_2\big) \frac{\partial q}{\partial e} \frac{\partial^2 q}{\partial e \partial a} - \frac{\partial q}{\partial a} E_1 \right] < 0 \tag{C.29}$$

Numerator N_{18}:

$$\frac{\partial \theta_L^i}{\partial d} = \frac{N_{18}}{D} < 0 \tag{C.30}$$

$$N_{18} = A_2 \big(1+\phi_1 q(e)\big) \frac{1-x}{(1-x-\theta_H^i)^2} \Big(\phi_1+\phi_2\Big) \left[-\frac{\partial K}{\partial \kappa} E_1 K_1 - \big(\omega_L^f (x-\theta_L^i) + \omega_H^f (1-x-\theta_H^i)\big)\big(\phi_1-\phi_2\big) \frac{\partial^2 K}{\partial \kappa \partial d} \right] < 0 \tag{C.31}$$

Numerator N_{28}:

$$\frac{\partial \theta_H^i}{\partial d} = \frac{N_{28}}{D} < 0 \tag{C.32}$$

$$N_{28} = A_1 \big(1+\phi_1 q(e)\big) \frac{x}{\big(x-\theta_H^i\big)^2} \left[-\frac{\partial K}{\partial d} E_1 K_1 + \Big(\omega_L^f (x-\theta_L^i) + \omega_H^f (1-x-\theta_H^i)\Big) \frac{\partial^2 K}{\partial \kappa \partial d} \frac{\partial K}{\partial \kappa} E_1 \right] < 0 \tag{C.33}$$

The impact of the taxes τ on θ_L^i and θ_H^i is positive. If there are higher taxes (contributions in the model) the informality is higher:

Numerator N_{19}:

$$\frac{\partial \theta_L^i}{\partial \tau} = \frac{N_{19}}{D} > 0 \tag{C.34}$$

$$N_{19} = A_2 \frac{1-x}{\big(1-x-\theta_H^i\big)^2} \frac{1+K(\kappa)}{1+\tau} \big(1+\phi_1 q(e)\big) E_1 K_1 > 0 \tag{C.35}$$

Numerator N_{29}:

$$\frac{\partial \theta_H^i}{\partial \tau} = \frac{N_{29}}{D} > 0 \tag{C.36}$$

$$N_{29} = A_1 \frac{x}{\left(x - \theta_L^i\right)^2} \frac{1 + K(\kappa)}{1 + \tau} \big(1 + \phi_1 q(e)\big) E_1 K_1 > 0 \tag{C.37}$$

The effect of the cost parameters (b_1 and b_2) on the informality is positive in all cases. If the enforcement and the benefits are more expensive, the informality goes up.

Numerator N_{110}:

$$\frac{\partial \theta_L^i}{\partial b_1} = \frac{N_{110}}{D} > 0 \tag{C.38}$$

$$N_{110} = -\frac{A_2\big(1 + K(\kappa)\big)\big(\phi_1 + \phi_2\big)}{1 - \phi_2 q(e)} \frac{1 - x}{\left(1 - x - \theta_H^i\right)^2} \frac{\partial^2 B}{\partial e \partial b_1} \frac{\partial q}{\partial e} K_1 < 0 \tag{C.39}$$

Numerator N_{210}:

$$\frac{\partial \theta_H^i}{\partial b_1} = \frac{N_{210}}{D} > 0 \tag{C.40}$$

$$N_{210} = -\frac{A_1\big(1 + K(\kappa)\big)\big(\phi_1 + \phi_2\big)}{1 - \phi_2 q(e)} \frac{x}{\left(x - \theta_L^i\right)^2} \frac{\partial^2 B}{\partial e \partial b_1} \frac{\partial q}{\partial e} K_1 < 0 \tag{C.41}$$

Numerator N_{111}:

$$\frac{\partial \theta_L^i}{\partial b_2} = \frac{N_{111}}{D} > 0 \tag{C.42}$$

$$N_{111} = -A_2 \frac{1 - x}{\left(1 - x - \theta_H^i\right)^2} \frac{\partial^2 B}{\partial \kappa \partial b_2} \frac{\partial K}{\partial \kappa} E_1 < 0 \tag{C.43}$$

Numerator N_{211}:

$$\frac{\partial \theta_H^i}{\partial b_2} = \frac{N_{211}}{D} > 0 \tag{C.44}$$

$$N_{211} = -A_1 \frac{x}{\left(x - \theta_L^i\right)^2} \frac{\partial^2 B}{\partial \kappa \partial b_2} \frac{\partial K}{\partial \kappa} E_1 < 0 \tag{C.45}$$

1.8.2 Comparative statics of e^*

The effect of ϕ_1 on the equilibrium enforcement is negative, if there are more fines the quality of the enforcement is lower:

Numerator N_{31}:

$$\frac{\partial e}{\partial \phi_1} = \frac{N_{31}}{D} < 0 \tag{C.46}$$

$$N_{31} = A_2\big(1+K(\kappa)\big)q(e)\frac{1-x}{(1-x-\theta_H^i)^2}(\phi_1-\phi_2)\frac{\partial q}{\partial e}\omega_L^i K_1 - \big(1-\phi_2 q(e)\big)\big(\omega_L^i\theta_L^i+\omega_H^i\theta_H^i\big)\frac{\partial q}{\partial e}\frac{\partial K}{\partial \kappa}\left(\frac{x}{x-\theta_L^i}A_1(\phi_1-\phi_2)\frac{\partial q}{\partial e}\omega_H^f + \frac{1-x}{1-x-\theta_H^i}A_2\frac{\partial K}{\partial \kappa}\omega_L^f\right) < 0 \tag{C.47}$$

When ϕ_2 is higher the sign of the effect on e^* is uncertain.

Numerator N_{32}:

$$\frac{\partial e}{\partial \phi_2} = \frac{N_{32}}{D} \quad \text{n.d.} \tag{C.48}$$

$$N_{32} = \big(1+\phi_1 q(e)\big)\Big[\frac{\partial K(\kappa)}{\partial \kappa}\Big]^2\Big(\frac{\partial q}{\partial e}\Big)\big(\omega_L^i\theta_L^i+\omega_H^i\theta_H^i\big)\left(\frac{A_1\omega_H^f x}{(x-\theta_L^i)^2}+\frac{A_2\omega_L^f(1-x)}{(1-x-\theta_H^i)^2}\right) \frac{1+K(\kappa)}{1-\phi_2 q(e)}\big(1+\phi_1 q(e)\big)\big(\phi_1-\phi_2\big)q(e)\Big(\frac{\partial q}{\partial e}\Big)K_1\left(\frac{A_1\omega_H^i x}{(x-\theta_L^i)^2}+\frac{A_2\omega_L^i(1-x)}{(1-x-\theta_H^i)^2}\right) \frac{1-x}{(1-x-\theta_H^i)^2}\frac{x}{(x-\theta_L^i)^2}\big(\omega_L^i\theta_L^i+\omega_H^i\theta_H^i\big)\Big(\frac{\partial q}{\partial e}\Big)K_1 \quad \text{n.d.} \tag{C.49}$$

The comparative statics of e^* with respect to δ_2 depend on B_2. The sign goes in the same direction as the sign of the change in informality.

Numerator N_{33}:

$$\begin{aligned} \frac{\partial e}{\partial \delta_2} &= \frac{N_{33}}{D} > 0 \quad if \quad B_2 > 1 \\ \frac{\partial e}{\partial \delta_2} &= \frac{N_{33}}{D} < 0 \quad if \quad B_2 < 1 \end{aligned} \tag{C.50}$$

$$N_{33} = -\frac{\ln B_2 B_2^{\frac{-1}{\delta_2+1}}}{\left(\delta_2+1\right)^2}\frac{1-x}{(1-x-\theta_H^i)^2}\left[\omega_L^i(\phi_1-\phi_2)\frac{\partial q}{\partial e}K_1\right] > 0 \quad if B_2 > 1$$
$$N_{33} = -\frac{\ln B_2 B_2^{\frac{-1}{\delta_2+1}}}{\left(\delta_2+1\right)^2}\frac{1-x}{(1-x-\theta_H^i)^2}\left[\omega_L^i(\phi_1-\phi_2)\frac{\partial q}{\partial e}K_1\right] < 0 \quad if B_2 < 1 \tag{C.51}$$

Comparative statics of e^* respect to ψ_2:

Numerator N_{34}:

$$\frac{\partial e}{\partial \psi_2} = \frac{N_{34}}{D} < 0 \tag{C.52}$$

$$N_{34} = \frac{C_2}{(1-\psi_2)^2}\frac{1-x}{(1-x-\theta_H^i)^2}\left[\omega_L^i(\phi_1-\phi_2)\frac{\partial q}{\partial e}K_1\right] < 0 \tag{C.53}$$

The comparative statics of e^* with respect to δ_1 depend on B_1. The sign goes in the same direction as the sign of the change in informality.

Numerator N_{35}:

$$\frac{\partial e}{\partial \delta_1} = \frac{N_{33}}{D} > 0 \quad if \quad B_1 > 1$$
$$\frac{\partial e}{\partial \delta_1} = \frac{N_{33}}{D} < 0 \quad if \quad B_1 < 1 \tag{C.54}$$

$$N_{35} = -\frac{\ln B_1 B_1^{\frac{-1}{\delta_1+1}}}{\left(\delta_1+1\right)^2}\frac{x}{(x-\theta_L^i)^2}\left[\omega_H^i(\phi_1-\phi_2)\frac{\partial q}{\partial e}K_1\right] > 0 \quad if \quad B_1 > 1$$
$$N_{35} = -\frac{\ln B_1 B_1^{\frac{-1}{\delta_1+1}}}{\left(\delta_1+1\right)^2}\frac{x}{(x-\theta_L^i)^2}\left[\omega_H^i(\phi_1-\phi_2)\frac{\partial q}{\partial e}K_1\right] < 0 \quad if \quad B_1 < 1 \tag{C.55}$$

As in the case of ψ_2, the effect on e^* of a positive change in ψ_1 is negative, when the informal share has lower weight in the production function, the equilibrium share is lower and also the level of enforcement.

Numerator N_{36}:

$$\frac{\partial e}{\partial \psi_2} = \frac{N_{34}}{D} < 0 \tag{C.56}$$

$$N_{36} = \frac{C_1}{(1-\psi_1)^2}\frac{x}{(x-\theta_L^i)^2}\left[\omega_H^i(\phi_1-\phi_2)\frac{\partial q}{\partial e}K_1\right] < 0 \tag{C.57}$$

The effect of the quality parameters is uncertain in the case of a, which is one in the enforcement function, and negative in the one of the benefits function (d).

Numerator N_{37}:

$$\frac{\partial e}{\partial a} = \frac{N_{37}}{D} \quad \text{n.d.} \tag{C.58}$$

$$\begin{aligned} N_{37} = &-\big(1+\phi_1 q(e)\big)\left(\omega_L^i\theta_L^i + \omega_H^i\theta_H^i\right)\big(\phi_1-\phi_2\big)\frac{\partial^2 q}{\partial e \partial a}\left[\frac{\partial K}{\partial \kappa}\right]^2\left(\frac{A_1\omega_H^f x}{(x-\theta_L^i)^2} + \frac{A_2\omega_L^f(1-x)}{(1-x-\theta_H^i)^2}\right) \\ &\frac{1+K(\kappa)}{1-\phi_2 q(e)}\big(\phi_1^2-\phi_2^2\big)\frac{\partial q}{\partial e}\frac{\partial q}{\partial a}K_1\left(\frac{A_1\omega_H^i x}{(x-\theta_L^i)^2} + \frac{A_2\omega_L^i(1-x)}{(1-x-\theta_H^i)^2}\right) \\ &- \frac{x}{(x-\theta_L^i)^2}\frac{1-x}{(1-x-\theta_H^i)^2}\left(\omega_L^i\theta_L^i + \omega_H^i\theta_H^i\right)\big(\phi_1-\phi_2\big)\frac{\partial^2 q}{\partial e\partial a}K_1 \quad \text{n.d.} \end{aligned} \tag{C.59}$$

Numerator N_{38}:

$$\frac{\partial e}{\partial d} = \frac{N_{38}}{D} < 0 \tag{C.60}$$

$$\begin{aligned} N_{38} = \left[\big(1+\phi_1 q(e)\big)\big(\phi_1-\phi_2\big)\frac{\partial K}{\partial \kappa}\frac{\partial q}{\partial e}\right]\left[\left(K_1 - \left(\omega_L^f(x-\theta_L^i) + \omega_H^f(1-x-\theta_H^i)\right)\frac{\partial^2 K}{\partial e \partial d}\right)\right. \\ \left.\left(\frac{A_1\omega_H^i x}{(x-\theta_L^i)^2} + \frac{A_2\omega_L^i(1-x)}{(1-x-\theta_H^i)^2}\right)\right] < 0 \end{aligned} \tag{C.61}$$

The effect of the taxes on the equilibrium enforcement is positive, there are more resources to spend and one of the ways to do it is through the level of enforcement.

Numerator N_{39}:

$$\frac{\partial e}{\partial \tau} = \frac{N_{39}}{D} > 0 \tag{C.62}$$

$$N_{39} = -\frac{1+K(\kappa)}{1+\tau}\big(1+\phi_1 q(e)\big)\big(\phi_1-\phi_2\big)\frac{\partial q}{\partial e}K_1\left[\frac{A_1 x\omega_H^i}{\big(x-\theta_L^i\big)^2} + \frac{A_2(1-x)\omega_L^i}{\big(1-x-\theta_H^i\big)^2}\right] > 0 \tag{C.63}$$

Comparative statics of e respect to b_1 and b_2:

Numerator N_{310}:

$$\frac{\partial e}{\partial b_1} = \frac{N_{310}}{D} \tag{C.64}$$

Comparative statics of e respect to b_1 and b_2:

Numerator N_{311}:

$$\frac{\partial e}{\partial b_2} = \frac{N_{311}}{D} \tag{C.65}$$

$$N_{311} = \big(1+\phi_1 q(e)\big)\big(\phi_1-\phi_2\big)\frac{\partial K}{\partial \kappa}\frac{\partial q}{\partial e}\frac{\partial B}{\partial \kappa \partial b_2}\left[\frac{A_1 x \omega_H^i}{\big(x-\theta_L^i\big)^2}+\frac{A_2(1-x)\omega_L^i}{\big(1-x-\theta_H^i\big)^2}\right] > 0 \tag{C.66}$$

1.8.3 Comparative statics of κ^*

The effect of ϕ_1 on the equilibrium level of benefits is positive, more fines produce more resources for the formal workers:

Numerator N_{41}:

$$\frac{\partial \kappa}{\partial \phi_1} = \frac{N_{41}}{D} > 0 \tag{C.67}$$

$$N_{41} = \left[A_1\omega_H^f \frac{x}{(x-\theta_L^i)^2} + A_2\omega_L^f \frac{1-x}{(1-x-\theta_H^i)^2}\right] * \left[\frac{1+K(\kappa)}{1-\phi_2 q(e)}(\phi_1+\phi_2)\big(\omega_L^i\theta_L^i+\omega_H^i\theta_H^i\big)\Big[\frac{\partial q}{\partial e}\Big]^2\frac{\partial K}{\partial \kappa} - \big(1+K(\kappa)\big)q(e)\frac{\partial K}{\partial \kappa}E_1\right] > 0 \tag{C.68}$$

Comparative statics of e respect to ϕ_2:

Numerator N_{42}:

$$\frac{\partial e}{\partial \phi_2} = \frac{N_{42}}{D} < 0 \tag{C.69}$$

$$N_{42} = -\frac{1+K(\kappa)}{1-\phi_2 q(e)}\big(\phi_1+\phi_2\big)\Big[\frac{\partial K(\kappa)}{\partial \kappa}\Big]^2\Big[\frac{\partial q}{\partial e}\Big]^2\big(\omega_L^i\theta_L^i+\omega_H^i\theta_H^i\big)\left(\frac{A_1\omega_H^f x}{(x-\theta_L^i)^2}+\frac{A_2\omega_L^f(1-x)}{(1-x-\theta_H^i)^2}\right)$$
$$\frac{1+K(\kappa)}{1-\phi_2 q(e)}\big(1+\phi_1 q(e)\big)q(e)\Big(\frac{\partial K}{\partial \kappa}\Big)E_1\left(\frac{A_1\omega_H^f x}{(x-\theta_L^i)^2}+\frac{A_2\omega_L^f(1-x)}{(1-x-\theta_H^i)^2}\right)<0 \tag{C.70}$$

The effect of δ_2 in the κ^* depend on B_2 as in the other variables, but the sign is the opposite than in the case of the informal shares and the enforcement.

Numerator N_{43}:

$$\begin{aligned}\frac{\partial \kappa}{\partial \delta_2} &= \frac{N_{43}}{D} < 0 \quad if \quad B_2 > 1\\ \frac{\partial \kappa}{\partial \delta_2} &= \frac{N_{43}}{D} > 0 \quad if \quad B_2 < 1\end{aligned} \tag{C.71}$$

$$\begin{aligned}N_{43} &= \frac{\ln B_2 B_2^{\frac{-1}{\delta_2+1}}}{\big(\delta_2+1\big)^2}\frac{1-x}{(1-x-\theta_H^i)^2}\omega_L^f\frac{\partial K}{\partial \kappa}E_1 < 0 \quad if \quad B_2 > 1\\ N_{43} &= \frac{\ln B_2 B_2^{\frac{-1}{\delta_2+1}}}{\big(\delta_2+1\big)^2}\frac{1-x}{(1-x-\theta_H^i)^2}\omega_L^f\frac{\partial K}{\partial \kappa}E_1 > 0 \quad if \quad B_2 < 1\end{aligned} \tag{C.72}$$

The effect of ψ_1 and ψ_2 is positive on κ. When the weight of the formal workers in the production function is higher the level of benefits in equilibrium goes in the same direction:

Numerator N_{44}:

$$\frac{\partial K}{\partial \psi_2} = \frac{N_{44}}{D} > 0 \tag{C.73}$$

$$N_{44} = -\frac{C_2}{(1-\psi_2)^2}\frac{1-x}{(1-x-\theta_H^i)^2}\omega_L^f\frac{\partial K}{\partial \kappa}E_1 > 0 \tag{C.74}$$

The effect of δ_1 is similar as in the case of δ_2:

Numerator N_{45}:

$$\begin{aligned} \frac{\partial \kappa}{\partial \delta_1} &= \frac{N_{45}}{D} < 0 \quad if \quad B_1 > 1 \\ \frac{\partial \kappa}{\partial \delta_1} &= \frac{N_{45}}{D} > 0 \quad if \quad B_1 < 1 \end{aligned} \tag{C.75}$$

$$\begin{aligned} N_{45} &= \frac{\ln B_1 B_1^{\frac{-1}{\delta_1+1}}}{\left(\delta_1+1\right)^2} \frac{x}{(x-\theta_L^i)^2} \omega_H^f \frac{\partial K}{\partial \kappa} E_1 < 0 \quad if \quad B_2 > 1 \\ N_{45} &= \frac{\ln B_1 B_1^{\frac{-1}{\delta_1+1}}}{\left(\delta_1+1\right)^2} \frac{x}{(x-\theta_L^i)^2} \omega_H^f \frac{\partial K}{\partial \kappa} E_1 > 0 \quad if \quad B_2 < 1 \end{aligned} \tag{C.76}$$

Comparative statics of κ respect to ψ_1:

Numerator N_{46}:

$$\frac{\partial K}{\partial \psi_1} = \frac{N_{46}}{D} > 0 \tag{C.77}$$

$$N_{46} = -\frac{C_1}{(1-\psi_1)^2} \frac{x}{(x-\theta_L^i)^2} \omega_H^f \frac{\partial K}{\partial \kappa} E_1 > 0 \tag{C.78}$$

The effect of the quality parameters as is expected is positive in the level of benefits in equilibrium:

Numerator N_{47}:

$$\frac{\partial \kappa}{\partial a} = \frac{N_{47}}{D} > 0 \tag{C.79}$$

$$\begin{aligned} N_{47} = &\left(\frac{A_1 \omega_H^f x}{(x-\theta_L^i)^2} + \frac{A_2 \omega_L^f (1-x)}{(1-x-\theta_H^i)^2} \right) \frac{1+K(\kappa)}{1-\phi_2 q(e)} \frac{\partial K}{\partial \kappa} * \\ &\left(\left(\omega_L^i \theta_L^i + \omega_H^i \theta_H^i\right) \left(\phi_1^2 \phi_2^2\right) \frac{\partial^2 q}{\partial e \partial a} \frac{\partial q}{\partial e} - \left(\phi_1 + \phi_2\right) \frac{\partial q}{\partial a} E_1 \right) > 0 \end{aligned} \tag{C.80}$$

Numerator N_{48}:

$$\frac{\partial \kappa}{\partial d} = \frac{N_{48}}{D} > 0 \tag{C.81}$$

$$N_{48} = \frac{1+K(\kappa)}{1-\phi_2 q(e)}\left(\omega_L^i\theta_L^i + \omega_H^i\theta_H^i\right)\frac{\partial^2 K}{\partial e \partial d}\left(\phi_1^2 - \phi_2^2\right)\left(\frac{A_1\omega_H^i x}{(x-\theta_L^i)^2} + \frac{A_2\omega_L^i(1-x)}{(1-x-\theta_H^i)^2}\right)$$
$$- \left(1+\phi_1 q(e)\right)\frac{\partial K}{\partial \kappa}\frac{\partial K}{\partial d}E_1\left(\frac{A_1\omega_H^f x}{(x-\theta_L^i)^2} + \frac{A_2\omega_L^f(1-x)}{(1-x-\theta_H^i)^2}\right) \tag{C.82}$$
$$- \frac{x}{(x-\theta_L^i)^2}\frac{1-x}{(1-x-\theta_H^i)^2}\left(\omega_L^i\theta_L^i + \omega_H^i\theta_H^i\right)\frac{\partial^2 K}{\partial e \partial d}E_1 > 0$$

The effect of taxes (τ) is negative on the benefits, there are fewer formal workers and the government in equilibrium spends more money on the enforcement task:

Numerator N_{49}:

$$\frac{\partial \kappa}{\partial \tau} = \frac{N_{49}}{D} < 0 \tag{C.83}$$

$$N_{49} = \frac{1+K(\kappa)}{1+\tau}\left(1+\phi_1 q(e)\right)\frac{\partial K}{\partial \kappa}E_1\left[\frac{A_1 x\omega_H^f}{\left(x-\theta_L^i\right)^2} + \frac{A_2(1-x)\omega_L^f}{\left(1-x-\theta_H^i\right)^2}\right] < 0 \tag{C.84}$$

Comparative statics of κ respect to b_1:

Numerator N_{410}:

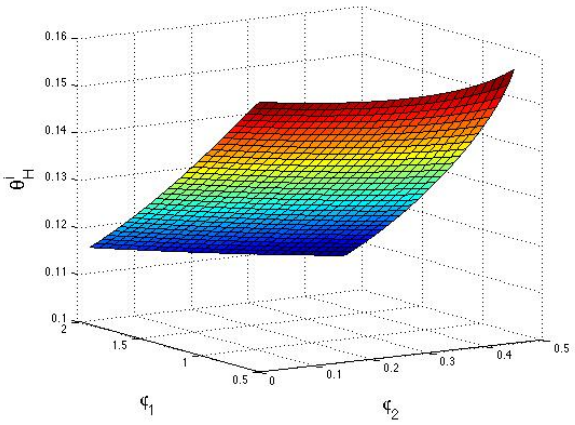
$$\frac{\partial \kappa}{\partial b_1} = \frac{N_{410}}{D} < 0 \tag{C.85}$$

$$N_{410} = -\frac{1+K(\kappa)}{1-\phi_2 q(e)}\left(\phi_1 + \phi 2\right)\frac{\partial K}{\partial \kappa}\frac{\partial q}{\partial e}\frac{\partial^2 B}{\partial e \partial b_1}\left[\frac{A_1 x\omega_H^i}{\left(x-\theta_L^i\right)^2} + \frac{A_2(1-x)\omega_L^f}{\left(1-x-\theta_H^i\right)^2}\right] < 0 \tag{C.86}$$

1.8.4 Comparative statics with respect to a couple of parameters

Figure C.1: Changes in the equilibrium with the fines (ϕ_1) and the informal wage losses (ϕ_2)

(a) θ_H^{i*}



(b) θ_L^{i*}

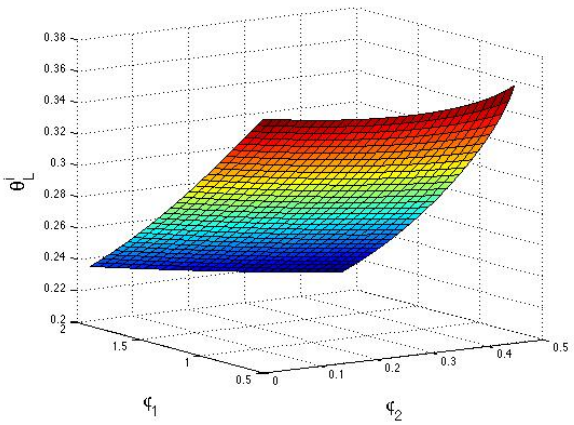
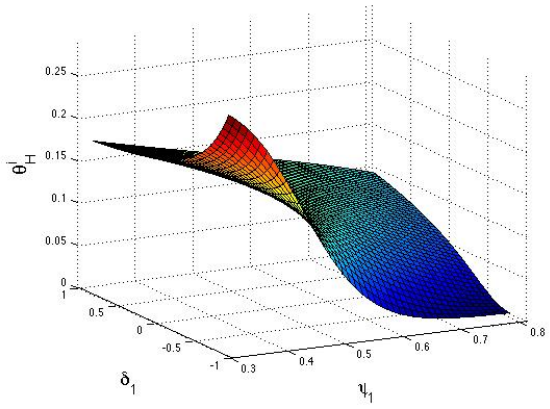
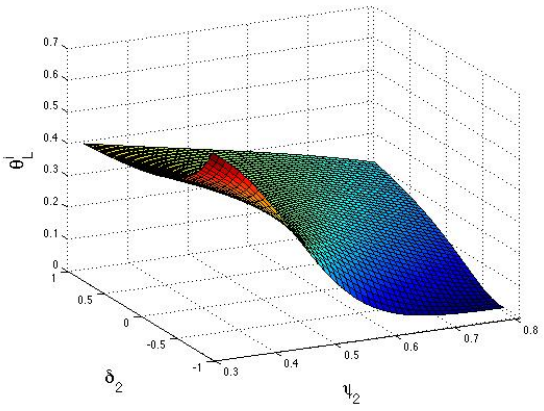


Figure C.2: Changes in the equilibrium with the level of substitutability and the shares (δ_j and ψ_j).

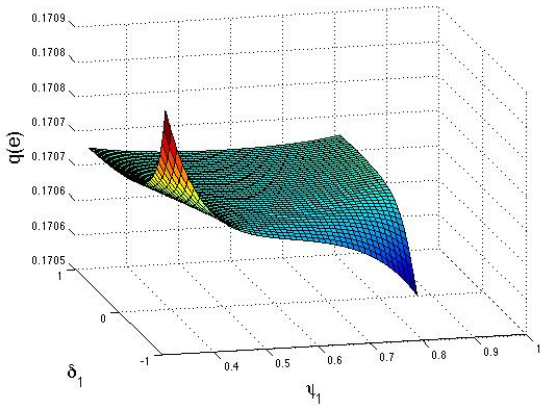
(a) θ_H^{i*} (δ_1,ψ_1)



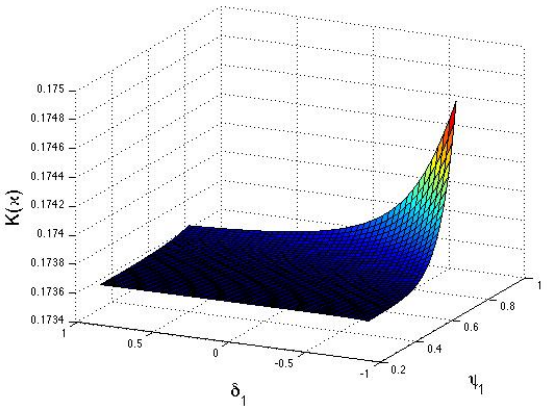
(b) θ_L^{i*}(δ_2,ψ_2)



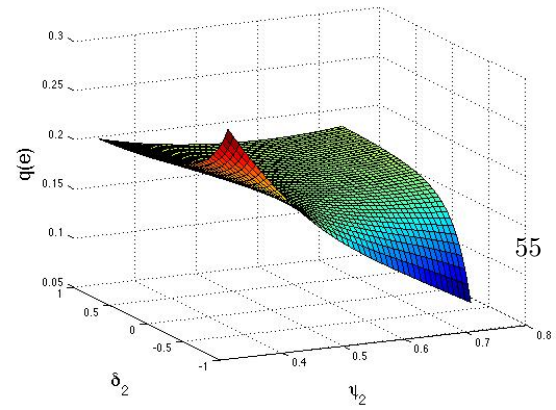
(c) $q(e)$ (δ_1,ψ_1)



(d) $K(\kappa)$ (δ_1,ψ_1)



(e) $q(e)$ (δ_2,ψ_2)



(f) $K(\kappa)$ (δ_2,ψ_2)

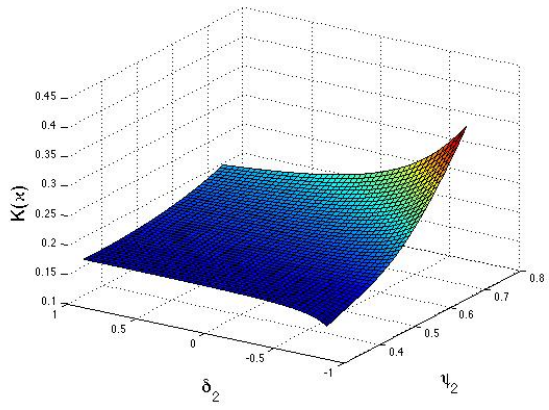
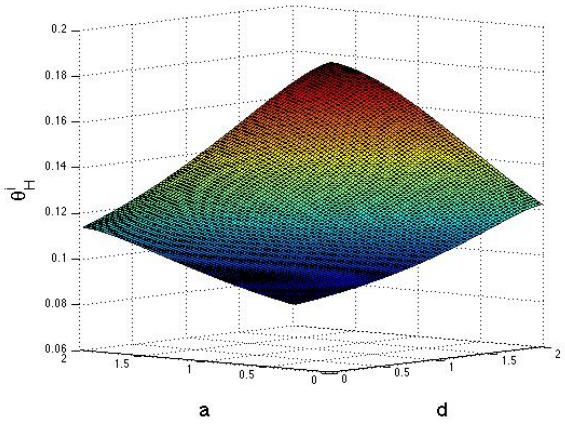
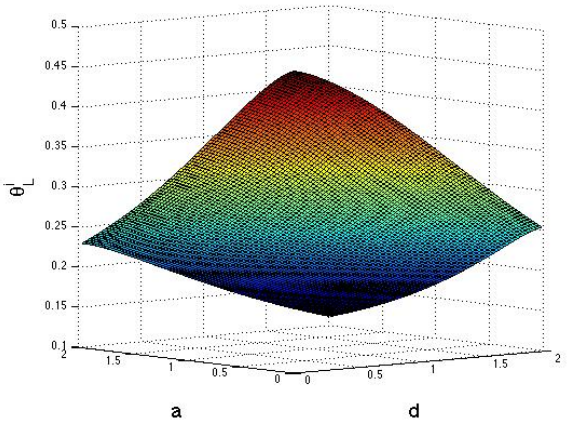


Figure C.3: Changes in the equilibrium with the quality parameters (a and d)

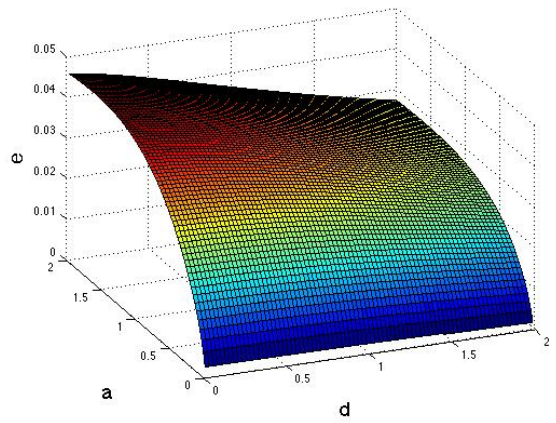
(a) θ_H^{i*}



(b) θ_L^{i*}



(c) e^*



(d) κ^*

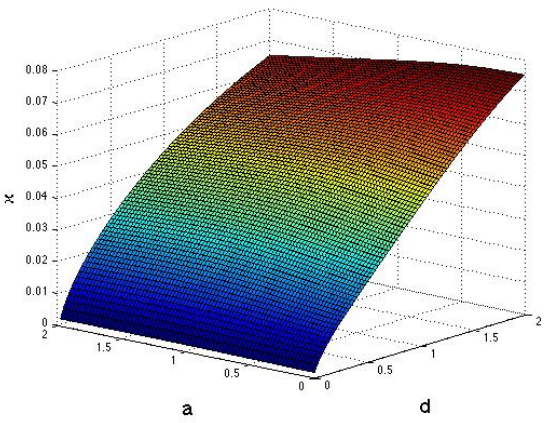
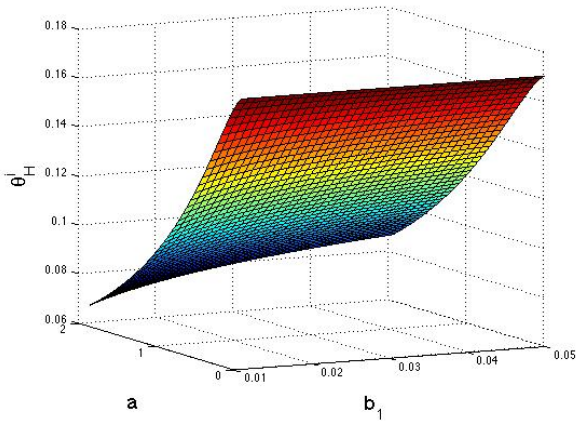
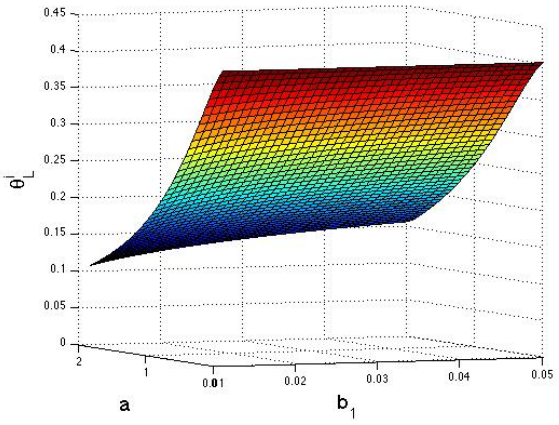


Figure C.4: Changes in the equilibrium with quality and cost of the enforcement (a) and (b_1)

(a) θ_H^{i*}



(b) θ_L^{i*}



(c) Enforcement

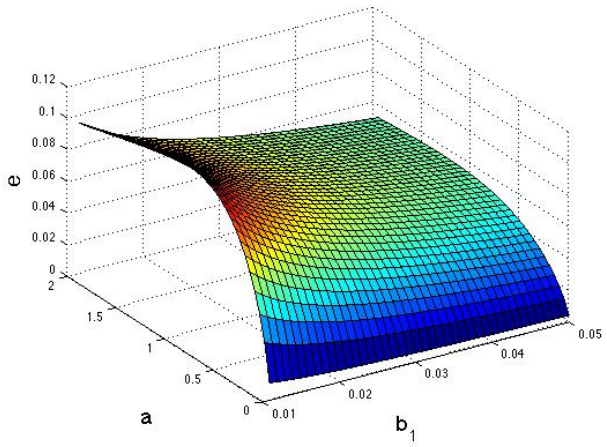
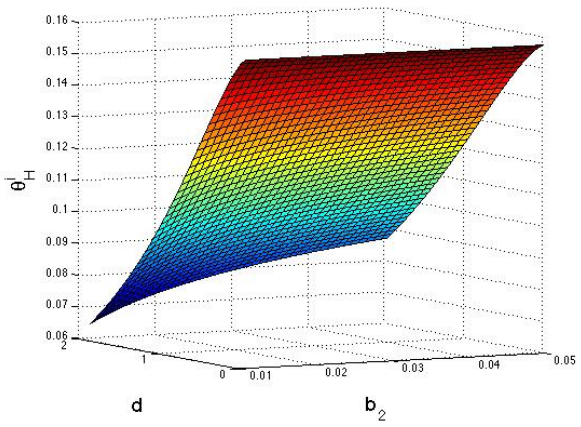
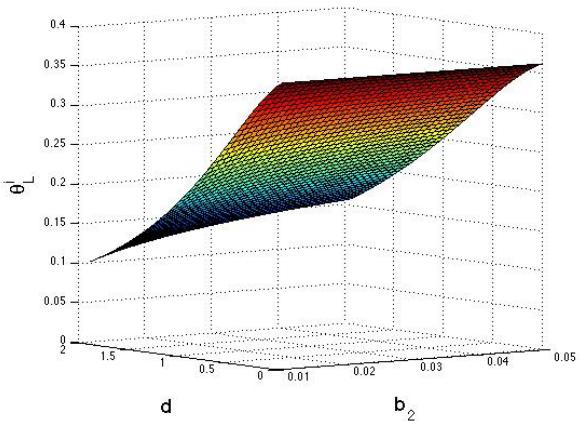


Figure C.5: Changes in the equilibrium with quality and cost of the benefits (d) and (b_2)

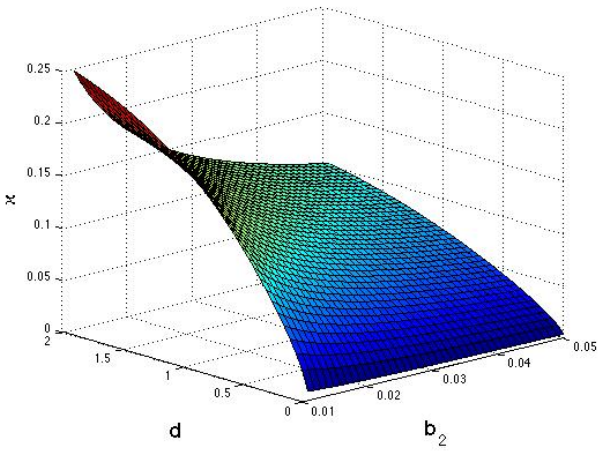
(a) θ_H^{i*}



(b) θ_L^{i*}



(c) Benefits



Chapter 2

Drop-out and enforcement during Two Transfer Programs

2.1 Introduction

The aim of this paper is to analyze the high school drop-out dynamic, specifically we focus on those teenagers[1] in families which are affected by income shocks because they participate in Conditional Cash Transfers (CCT) Programs, and how these behavior changes can be modified by the enforcement level. We develop a structural discrete choice model, where the decisions are jointly taken by the teenagers and their parents. To estimate the model parameters, we use data from two programs designed and carried out in Uruguay in the last decade, the Social Assistance National Plan to the Social Emergency (Plan de Asistencia Nacional a la Emergencia, hereinafter PANES) and the Family Allowances (Asignaciones familiares hereinafter AFAM). These program have requirements for the participants and one of those is school attendance for those individuals under 18. However, the level of enforcement in the programs is different not only because of the program design, but also over time.

In recent years the share of teenagers who drop-out of high-school and do not enter the labor market in Uruguay has been a focal point for policy makers and the academia. According to the ILO (2013), in Uruguay which is in an intermediate level in Latin America, one in five individuals aged between 14 and 19 do not study nor work. In the same report, they disentangle the activities that these individuals do, discriminating those who are engaged in home production, those who are jobseekers, and those engaged in other (inactive) activities . The most worrying feature of this figure in Uruguay and Paraguay is the high proportion, about 50%, of those individuals who answer that they spend their time in other (inactive) activities. Combining both statistics shows that Uruguay is the country in the region where the participation of the individuals between 14 and 24 who

[0]This chapter is coauthored with Gonzalo Salas.

[1]In this paper we consider as teenagers individuals between 12 and 18 years old.

neither study nor work is 10% of the total. Additionally, UM/CIEA (2013) indicates the share of age between 15 and 29 in the first quintile of income who neither study nor work, doubled in the last 5 years.

High-school drop-out can be measured with those teenagers who start to attend at the beginning of the academic years and then quit high school. If we consider those students who have at least 50 absences during the academic year and are not enrolled in the educative system in the next academic year, the average rate of drop-out in compulsory high school has been around 5% yearly over the last decade[2]. Note that, this is a lower bound given the threshold is high, considering both 50 absences and the non-enrollment.

Moreover, if we analyze the problem by socio-economic stratus the differences are dramatic. The rate of attendance of those teenagers who are in the fifth quintile of income is around 85% in compulsory and non compulsory education, but in the first quintile it is only 60% for compulsory and around only 25% in non compulsory education (Figure 2.1). This issue shows first a severe inequality problem, and second the incapacity to jointly improve the relative development with the GDP growth. Twenty years ago, Uruguay was at the top of educational performance in Latin America, but after the severe economic and social crisis in 2002-2003 the country has not being able to back that performance even though the rate of growth in the decade was the highest in its history.

The nature of drop-out is essentially dynamic. Poor educational performance, i.e. low Grade Point Average (GPA), in the past has increased its probability (Alexander et al. (2001); Griffin (2002); Christle et al. (2002)). In this specific process there are two types of incentives which play a determinant role: i) individual incentives, poor performances can generate frustration in the individual (Finn (1989)) and can reduce how enjoyable it is to be in school (Stinebrickner and Stinebrickner (2013)); and ii) household incentive through the aspirations of parents, that is the child's educational performance builds parents' incentives. Because they visualize bad signals emitted by their offspring's outcome, they stop investing in education (Li and Mumford (2009)). This investment in education can operate either through the time that parents spend with their children in formative activities such as reading or homework, or encouraging their children to do it (Boca et al. (2012)).

The decision to participate in the education system depends on both parents' and

[2]The drop-out in the first three years is around 4% and 7% in the 4th one. Source: http://www.anep.edu.uy/observatorio/.

teenager's utilities. When the parents' utility is low because of poor educational performance, they can be compensated by more income if their offspring participates in the labor market. In the case of teenagers, their utility depends on leisure time and the time spent on alternative activities (school attendance or work). We assume that the utility that is extracted from attending school depends not only on the GPA but also on course achievement.

In their seminal paper Eckstein and Wolpin (1999) develop and estimate a structural model of work decision and high school attendance. They exploit the NLSY79[3] to know who drops out and when they do so. They found that those who work contemporaneously while they attend high school have lower levels in their school performance. When they analyze some policy experiments they assess some measures, such as work prohibition which has had some limited success in improving school outcomes. In our paper, we deal with a particular group of teenagers who are at the bottom of the income distribution and a significant share of them neither study nor work, which introduces a particular feature into our model.

Stinebrickner and Stinebrickner (2013) estimate a structural dynamic model to understand and quantify the different channels from which the college student drops out. They point out the role of GPA performance in this decision. This paper gives us many insights into the dynamic of the GPA, but the nature of the decision is quite different given that they studied adults and we are working with teenagers.

The economic and financial crisis in Uruguay in 2002 generated a high increase in unemployment and poverty. For this reason, in 2005 the PANES program was implemented. A fixed cash transfer was directed to the household regardless of the number of members. The target population of this program was the first quintile of the poorest population. Among the required conditions was attendance at school, although there is evidence of the low level of enforcement and compliance of the requirements (Labat (2012)). In December 2007, this program was ended. From the beginning it had been proposed as transitory and the families with children under 18 years were integrated into the AFAM program. This last program is also a CCT, with similar requirements, but it is part of the Social Protection System. In this case, the target population is to cover all poor households with children under 18. The amount of the cash transfer depends on the number of children and if they attend elementary or high school. This latter program include all the PANES

[3]The National Longitudinal Survey of Youth 1979.

beneficiaries with children under 18. The income threshold is higher than in PANES, and the amount of the transfer is similar but in the case of AFAM, it obviously has a bigger dispersion by household than AFAM.

The CCTs programs operate on the probability of drop-outs by two mechanisms. First, in a direct way, due to the fact that one of the conditions to participate in these programs refers to school attendance. Second, the programs indirectly generate behavioral changes based on a variation of incentives, decreasing the investment required to study or the opportunity cost of studying in relation to labor activities.

Todd and Wolpin (2006) analyze the effect of a transfer program PROGRESA in Mexico on child schooling and fertility. They develop a dynamic behavioral model where the parents first, and then the teenagers decide either to work or attend school and fertility behavior given the existence of a transfer program. Additionally, they perform some contrafactual policy alternatives and propose a different scheme which leads to better school performance. Attanasio et al. (2010) also use a structural model to evaluate the PROGRESA in Mexico. They exploit a randomized experiment to assess where the program is more effective and at which points it could be improved. Our paper goes one step further: we work with two transfer programs and we analyze how enforcement plays a role in school participation. Finally, we also include a grade dynamic in the model.

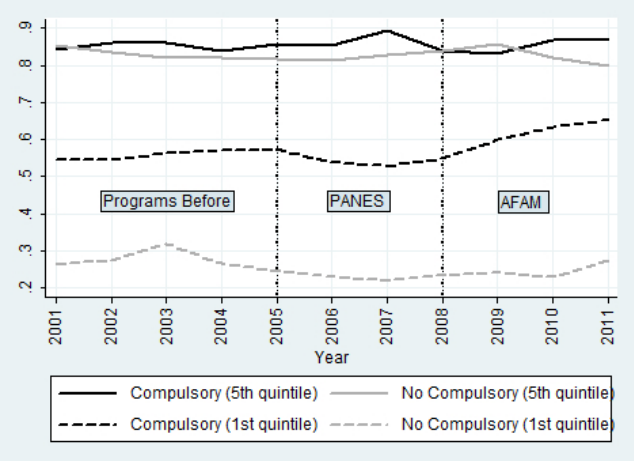


Figure 2.1: Rate of High School Attendance in the first and the fifth quintile and year. Source: Continuos Household Survey.

Enforcement is a concept that has been gaining a crucial role in public economic literature. It involves not only the resources that the government invests to carry out the programs, but also the individual perception about the quality and its efficiency. The individual enforcement perception and their externalities are introduced by Alm et al. (2009) in a paper about tax compliance in a lab experiment, and they were studied by Rincke and Traxler (2011) in an empirical paper about TV licenses in Austria. They identified enforcement spillovers where the activity of inspectors leads on average to an unsolicited registration for every three effectively enforced registrations.

Kaufmann et al. (2012) assess the enforcement relevance in CCT programs focused on income requirement with data from the *Bolsa familia* Program in Brazil. These authors found that people learn about the program enforcement not only with their own experience, but also with their peer experience. They found that the changes in behavior depend on public and private signals about the enforcement quality, and this feature is a key point in the program's effectiveness.

On comparing both programs, there are some similarities and many differences as is shown in Table 2.1. Both programs target the beneficiaries through a Baseline Poverty Score (Índice de Carencias Críticas, ICC in spanish) and the formal income per capita of the household. The ICC identifies the probability that the household is vulnerable, then the households above a threshold are eligible for the program. Both PANES and AFAM have the household as an objective, but the AFAM has a wider coverage. Both programs use different thresholds with PANES targeting the first quintile of the poorest population and AFAM putting the emphasis on all the poor children

The main differences are the amount of the transfer and the enforcement mechanism. The transfer in PANES is the same amount for all the households, a lump sum one. Conversely, the AFAM transfer is linked with the number of household members and the educative achievement. The first child of the household receives an amount, which is multiplied by 0.6 for the other younger child. If the children attend high-school they receive 30% more as a bonus.

The enforcement in both programs is different but neither is efficient in regulating the requirements (specifically education and health), given they are not a main concern for the policy maker. The enforcement is based on the individual perception that if they not hold the requirement they will lose the transfer. In terms of perception, PANES enforcement is higher than that of AFAM, because the probability of losing the entire transfer depends on

any single children. Conversely, in AFAM if one of the children drops out of the education system, the household will lose only the part of the transfer that corresponds to that child.

However, the enforcement perception can be understood as higher in the AFAM case, because when the children start high-school the family has to present the enrollment certificate in order to receive the 30% bonus. In 2013, the government checked in April if the children were enrolled in the educative system, and in September checked again the number of days that they effectively attend school. This change led to the suspension of some individuals from the program and could spillover to other families through the enforcement perception.

Table 2.1: Program design.

	PANES	AFAM
Targeting mechanism	ICC & Formal income per capita	ICC & Formal income per capita
Population	First quintile of poor households	Poor households with children
Transfer	Lump sum per household	Per children but decreasing with the number of children in the household and an extra to be in high-school.
Enforcement	Weak regulation. Perception based. The household loses the transfer if one member does not meet the requirements.	Weak regulation. Perception based. The household loses the part of the transfer of the member that does not meet the requirements. Regulation when the children enter high school. In 2013 strong monitoring.

In 2013, due to the fact the authorities increased the regulation, in April they cancelled 26,000 households (6% of the total) because the children were not enrolled in the school system, and in November they cancelled 10,500 more because the children did not attend enough days during the year. Among these transfer cancellations 40% were for children who should have attended high-school[4].

The income requirement has been monitored in both PANES and AFAM programs, but only the formal income which is registered through the labor records. All those incomes that the families have from informal jobs cannot be monitored by the government. Around

[4]In 2014, the enforcement agency continues with this policy.

5% of the PANES and AFAM beneficiaries exit from the programs because they are above the formal income threshold. The enforcement of this requirement is widely known and enters into the decision function of the individual[5].

Our paper analyzes **the dynamic of drop-out in compulsory and non compulsory education and, how it is affected by the shocks of income and the attendance enforcement when the households have access to one (or both) of the CCT programs.** We analyze individuals who go through the high school when there are two different programs, PANES and AFAM, which have the same objective of encouraging school participation but the enforcement perception is different.

We will focus on three points which are not analyzed in the literature: first, how is the utility formation for those who neither study nor work; second, what is the role of the time of home production, and finally we analyze how the CCT program is designed, particularly how is the function that determines the loss of the transfer and the level of enforcement that the government agency applies.

The rest of the paper is organized as follows. In section 2 we describe the data bases and the main descriptive statistics. In section 3 the model is developed. In section 4 the estimation strategy is presented. In section 5 we present the results and in section 6 we perform some policy experiments. Finally section 7 concludes.

2.2 The Data and Descriptive Statistics

Educational performance is a heterogeneous phenomenon, richer teenagers attend more classes than poorer ones, and even in compulsory education the difference is quite significant (see Figure 2.1). Around 85% of the richer teenagers attend compulsory and non-compulsory high-school, and around 60% of poorer teenagers attend compulsory and 25% non-compulsory high-school. The effect of the crisis can be seen in the decrease in attendance in the poorest ones; although this process did not stop in 2007 even though the PANES had been implemented. In 2011, when the AFAM was in progress, the attendance increased significantly, even when demand for labor was the highest in the history of this country.

This paper focused on the poorest population which is around the threshold determined to participate in the CCT programs. The information used in this paper comes from

[5]Similar features are identified by the World-Bank (2010) for Colombia.

administrative records and surveys that can be combined using the national ID number of the person. They are: the follow-up survey of PANES (FSP) and the high-school education record (SER). The PANES is a transient program that started in April 2005 and ended in December 2007. The most important component was a lump sum transfer[6] which is independent of the number of household members[7]. The target population of this program was the first quintile of the poorest households.

The FSP consists of data collected as part of the evaluation of the PANES program. We have two waves of this survey. The first wave is primarily from 2006, although part of it corresponds to 2007, and the second one corresponds to 2008. In this follow-up survey, it is possible to identify the beneficiaries of the program (treatment group) and those who applied but were not selected (control group). The beneficiary selection criterion arises from the ICC. This survey considers only the population that is around the cutoff that identifies the treated and untreated groups.

The AFAM transfer depends on the number of children in the household and the educative level. The amount for children depends on whether they are in high-school (30% bonus) and for the younger members of the household to the transfers a factor of 0.6 is applied.

In the same fashion as FSP, the Follow-up survey for AFAM (FSA) is an instrument used to evaluate the AFAM program. In this case, we have only one wave in 2011. The criteria allow us to identify the treated and control population in a similar way to the FSP, through the ICC and formal per capita income threshold.

To complement the FSP and FSA data, they are combined with information from SER, which contains data on the educational performance of students in secondary education, the Grade Point Average (GPA). This cycle starts at 12 years of age, after 6 years of primary education. This education stage is divided into two cycles, the first three years correspond to the basic cycle (compulsory education) and the last three to the advanced cycle (non compulsory high-school)[8]. Additionally, we estimate the home production time with the Use of Time Survey carried out in 2008 by the National Statistics Institute[9].

In Table 2.2, we show the mean and standard deviation of the main variables during the

[6]In spanish was called: Ingreso ciudadano.

[7]In addition, households with children received a food card (in-kind transfer) where amount depended on the number of children in the household.

[8]In spanish are called: Ciclo básico y bachillerato diversificado respectively.

[9]The result of the estimation model is presented in the Table D.4.

period of both CCT programs. In the first panel, the data about PANES can be observed. We consider 3090 observations of 12 to 18 year old individuals. Of this population 75% attended the school system. However only 55% attended compulsory high-school school (1716 observations) and 44% attended non-compulsory high-school (1362 observations). Of these cases we can only locate 707 students in the SER due to the absence of an ID. We do not observe significant changes in the distribution of variables as a consequence of the missing cases. In our sample 70% of the population were treated, nearly 65% carried out home production and 6% were working. Specific information about educational performance shows that 35% fail the course that they attend (obtain an F) and only 20% obtained a GPA of A. Finally, less than 9% attended 5th and 6th grade of high-school.

In the second panel we show data about the AFAM period. There, 82% attended formal education (7 points more than in the PANES). We consider 2796 individuals between 12 and 18 years old, and we have high-school records for 952 of them. The age and the treated population is similar to PANES. The estimation of home production is also quite similar. However, the AFAM population work less than the PANES one, 3 points less.

We define four states with the combination of studying and working choices. In Table 2.3, we present the distribution of hours worked and home production by age. Furthermore, we show the distribution of the states that are of our interest, which are teenagers who only study (sn), those who study and work (sw), those who neither study nor work (nn), and those who only work (nw). In this case we observe the number of teenagers who only study decreases significantly with age, but the trend is increasing for those who neither attend school nor work. Additionally, the percentage of those who study and work is always less than 10%. In the case of hours worked we note that it increases with age as expected. The increase in hours allocated to home production presents an irregular trend.

Comparing both programs, during AFAM there are more teenagers studying and not working and this is because of the decrease in those who neither study nor work.

The distribution of GPA by age and grade is presented in Table 2.4. The grade performance is worse when students attend higher level courses and with their age. About 64% of those older than 16 years old and 72% of those enrolled in 5th and 6th grade fail the course. This difference is due to the fact that students are enrolled in lower courses that would correspond to their age because of repeated fails. The percentage that obtains the best GPA is constant between first and fourth grade (20%), and decreases to only 10% in the last two grades.

				PANES					
		FSP		HS attendance (FSP)			FSP and SER		
	Obs.	Mean	S.D	Obs.	Mean	S.D	Obs.	Mean	S.D
Age 12-18									
Attendance	3090	0.746	0.435						
Age	3093	14.83	2.014	1330	14.60	1.769	707	14.11	1.566
Treatment	3093	0.701	0.458	1330	0.689	0.463	707	0.680	0.466
Home Production									
0	3079	0.333	0.471	1322	0.332	0.471	703	0.354	0.478
0-10	3079	0.315	0.464	1322	0.378	0.485	703	0.404	0.491
> 10	3079	0.351	0.477	1322	0.290	0.453	703	0.242	0.428
GPA									
F							707	0.349	0.477
B							707	0.444	0.497
A							707	0.206	0.405
Grade									
1-2							707	0.584	0.493
3-4							707	0.328	0.470
5-6							707	0.088	0.283
Age 14-18									
Hours									
0	2093	0.823	0.381	903	0.905	0.294	423	0.941	0.236
0-15	2093	0.062	0.241	903	0.041	0.198	423	0.031	0.173
> 15	2093	0.097	0.296	903	0.043	0.203	423	0.021	0.144
				AFAM					
		FSA		HS attendance (FSA)			FSA and SER		
	Obs.	Mean	S.D	Obs.	Mean	S.D	Obs.	Mean	S.D
Age 12-18									
Attendance	2796	0.821	0.382						
Age	2936	14.93	1.0	1555	14.82	1.72	952	14.43	1.47
Treatment	2936	0.731	0,443	1555	0.692	0.462	952	0.721	0.448
Home Production									
0	2641	0.141	0.348	1484	0.127	0.333	917	0.154	0.361
0-10	2641	0.541	0.498	1484	0.582	0.493	917	0.581	0.493
> 10	2641	0.318	0.466	1484	0.291	0.454	917	0.265	0.441
Grade									
1-2							952	0.574	0.494
3-4							952	0.425	0.494
Age 14-18									
Hours									
0	1901	0.868	0.338	1116	0.945	0.227	644	0.973	0.160
0-15	1901	0.03	0.18 0	1116	0.02	0.142	644	0.01	0.103
> 15	1901	0.10	0.297	1116	0.04	0.181	644	0.016	0.123

Table 2.2: Descriptive Statistics. Source: FSP, FSA and SER.

PANES										
	State				Hours Worked			Home Production		
	sn	*sw*	*nn*	*nw*	0	0-15	> 15	0	0-10	> 10
12-13	0.82	-.-	0.18	-.-	-.-	-.-	-.-	0.44	0.49	0.08
14-15	0.77	0.05	0.16	0.03	0.93	0.04	0.03	0.02	0.60	0.38
> 16	0.51	0.06	0.28	0.15	0.80	0.08	0.12	0.12	0.48	0.40
AFAM										
	State				Hours Worked			Home Production		
	sn	*sw*	*nn*	*nw*	0	0-15	> 15	0	0-10	> 10
12-13	0.96	-.-	0.04	-.-	-.-	-.-	-.-	0.40	0.52	0.08
14-15	0.86	0.02	0.10	0.03	0.95	0.02	0.03	0.08	0.58	0.34
> 16	0.61	0.07	0.20	0.12	0.81	0.04	0.15	0.00	0.52	0.48

Table 2.3: Decisions by group of age. Source: FSP, FSA and SER.

GPA	Age			Grade		
	12-13	14-15	> 16	1-2	3-4	5-6
F	0.22	0.40	0.64	0.34	0.33	0.72
B	0.50	0.43	0.26	0.44	0.46	0.18
A	0.28	0.17	0.10	0.22	0.21	0.10
Total	100.0	100.0	100.0	100.0	100.0	100.0

Table 2.4: The grades distribution by age group. Source: FSP and SER.

Finally, the transition rates between states in consecutive years are shown in Table 2.5. The state *sn* is more stable than the others, about one third of the population is only studying and remains there the next year. In any other case the proportion of the population in the same state exceeds 10%. The largest movements occur to study exclusively from all the other states. Of those *sw* in $t-1$, the next year about 60% stop working and continue studying, this percentage is 28% and 43% for the case of *nn* and *sw*, respectively.

2.3 Model

We develop a dynamic model of sequential decisions under uncertainty which is based on the basic model of the seminal paper of Eckstein and Wolpin (1999). Household utility depends on the time allocation of the teenager, whether he attends school, produces at home, works in the market or enjoys leisure. Additionally, that allocation determines if the household receives (or continues receiving) the CCT. Here we consider the utility that the teenager brings to the household weighting the utility that the teenager directly enjoys (U_{ch}), and the utility that the parents (U_p) enjoy through the teenager's time allocation.

	sn_{t-1}	sw_{t-1}	nn_{t-1}	nw_{t-1}	Total
sn_t	54.6	3.1	2.3	5.0	65.0
sw_t	4.4	0.6	0.4	0.6	5.9
nn_t	10.7	0.9	4.2	3.2	19.1
nw_t	4.9	0.7	1.5	2.8	10.0
Total	74.6	5.2	8.5	11.7	100.0

State Distribution of t on t-1

	sn_{t-1}	sw_{t-1}	nn_{t-1}	nw_{t-1}
sn_t	73.2	58.4	27.8	42.9
sw_t	5.9	10.9	4.3	5.4
nn_t	14.3	16.8	50.0	27.7
nw_t	6.6	13.9	17.9	24.1
Total	100.0	100.0	100.0	100.0

Table 2.5: Transitions of states. Source: FSP, FSA and SER.

The weight (γ_t) depends on the age of the teenager, if the age is below 14 the parent's weight is relatively higher than when they are over 14. The teenager values school attendance, market work and leisure time (total time minus the hours of market work and home production). The parents value school attendance, market work and home production.

$$U_t = \gamma_t U_{ch,t} + (1 - \gamma_t) U_{p,t} \tag{2.1}$$

This utility function could be thought of as the result of a bargaining process between teenager and parents about the teenager's time allocation where the bargaining power changes with the teenager's age. In the literature of family economics this formalization is used in the decision making of couples (Browning et al. (2014)), not of teenagers.

Figure 2.2 shows the choices that the household can take. The decision is how to split the time between school attendance, home production, leisure and market work, when they are legally able to work.

Given the total hours available L_1 (73 hours per week[10]) for those who attend school, L_2 (98 hours per week[11]) for those who do not attend the rewards in each situation $k = \{sn, sw, nn, nw\}$ depend on the value of attending school (b^s), the value of leisure (b^n), the value of working in the market (ωh^w), and the value of home production ($b^{hp} h^{hp}$). The utility is a weighted function of the teenager and parents' utility function.

[10]This computation is the result of considering that they have 14 hours available per day (after considering sleep, food and clean time) minus 25 weekly hours to attend school and study.

[11]This computation is the result of considering that they have14 hours available per day.

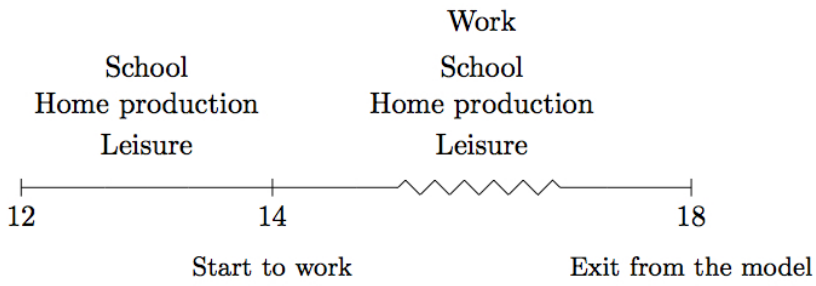


Figure 2.2: Timeline of the individual in the model by age (t)

The value when the teenager attends school and does not work (U^{sn}) includes the value of leisure, the value of studying for both, teenagers and parents; and the value of home production in the case of the utility of the parents. The value of studying and working (U^{sw}) includes the rewards of working (ωh^w), which is split between the teenager and their parents. The value of neither studying nor working (U^{nn}) includes only leisure and home production. The value of not studying and working includes home production and rewards of working (U^{nw}). Finally, the value of the CCT is included in the parents utility function, and it is multiplied by the enforcement parameter η, if the teenagers are not attending formal education. This parameter summarizes the enforcement spillover of the program, e.g. the perceived probability of being monitored. The transfer depends on the program and whether the family is above the formal income, insofar that income enforcement is perfect.

Teenager utility comes from leisure, school attendance and work in the market:

$$\begin{aligned} U^{sn}_{ch,t} &= B^n_{1,t}\left(L_1 - h^{hp}_t\right) + B^s_{1,t} \\ U^{sw}_{ch,t} &= B^n_{1,t}\left(L_1 - h^w_t - h^{hp}_t\right) + B^s_{1,t} + \omega_t h^w_t \\ U^{nn}_{ch,t} &= B^n_{1,t}\left(L_2 - h^{hp}_t\right) \\ U^{nw}_{ch,t} &= B^n_{1,t}\left(L_2 - h^w_t - h^{hp}_t\right) + \omega_t h^w_t \end{aligned} \tag{2.2}$$

Parent utility comes from school attendance, time allocated in home production (h^{hp}_t),

time working in the market (h_t^w) and the CCT (T_t):

$$\begin{aligned} U_{p,t}^{sn} &= B_{2,t}^{s} + B^{hp}(h_t^{hp}) + T_t \\ U_{p,t}^{sw} &= B_{2,t}^{s} + B^{hp}(h_t^{hp}) + \omega_t h_t^{w} + T_t \\ U_{p,t}^{nn} &= B^{hp}(h_t^{hp}) + \big(1-\eta\big)T_t \\ U_{p,t}^{nw} &= B^{hp}(h_t^{hp}) + \omega_t h_t^{w} + \big(1-\eta\big)T_t \end{aligned} \tag{2.3}$$

The value of the leisure for the teenager depends positively on the age (t), and it is convex in the hours that they do not spend working in the market, on home production, or in formal education. The value depend also on a set of parameters (b_{1t}^n, b_2^n and b_{11}^n).

$$B_{1t}^{n}(h) = b_{1t}^{n}(h^n)^{b_2^n} + b_{11}^{n} \tag{2.4}$$

The reward in school depends on the grades in the last period (gpa_{t-1}), the level of education achieved (E_{t-1}), the hours spent working in the market (h^{hw}) and on home production (h^{hp}) and two parameters (b_1^s and b_2^s), one for the teenager and other for the parents.

$$B_{1t}^{s} = b_1^{s}\left(gpa_{t-1}, E_{t-1}, h^{hp}, h^{hw}\right) + \epsilon_t^{s} \tag{2.5}$$

$$B_{2t}^{s} = b_2^{s}\left(gpa_{t-1}, E_{t-1}\right) + \epsilon_t^{s} \tag{2.6}$$

The enforcement parameters are different for the two CCT programs (η_{PANES} and η_{AFAM}):

$$\eta \in [0,1] \quad \text{if} \quad \eta = \begin{cases} 0 & \text{if enforcement does not exist} \\ 1 & \text{if full enforcement} \end{cases}$$

The grades follow an ordered logit process which depends on the age, the lag of grades, the work hours, the home production hours and the CCT. The grades can take three values

(A, B and F) the lower one means that the student fails the course.

$$\begin{aligned} gpa^* &= X\beta + e \qquad e/X \sim N(0,1) \\ gpa_t &= F \quad \text{if} \quad gpa^* < \nu_1 \\ gpa_t &= B \quad \text{if} \quad \nu_1 < gpa^* < \nu_2 \\ gpa_t &= A \quad \text{if} \quad gpa^* > \nu_2 \end{aligned}$$

The CCT can be received by a household where the teenager attends school in the first period. Then, in the following periods the household can continue receiving the CCT which depends on school attendance, an income shock and the government enforcement. The probability of losing the income of CCT programs by a formal income shock[12] are p_1 if the student is not working, and p_2 for those who are working. These percentages are estimated in two groups: for those who are between 12 and 14, and for those between 15 and 17 years old.

$$\begin{aligned} P(CCT = 0/CCT = 1, nw) = p_1(t) \\ P(CCT = 0/CCT = 1, w) = p_2(t) \end{aligned} \tag{2.7}$$

The reward of the home production depends on parameters b_t^{hp}, the age (t) and the hours on home production(h^{hp}):

$$B_t^{hp}(h) = b_{1t}^{hp} * t * (h^{hp})^{b_2^{hp}} \tag{2.8}$$

The wage in each moment is determined by the age (t), the school attendance (At_t), the level of education achieved (CS_{t-1}[13], B_{t-1}[14]), the time spent on home production (h_t^{hp}) and whether the household receives the CCT transfer. The estimation is done in two steps, in the first step we estimate the probability of being working, and in the second step including the Mills ratio to correct the selection bias.

$$\ln \omega_t = \beta_0 + \beta_1(t) + \beta_2 At_t + \beta_3 CS_{t-1} + \beta_4 B_{t-1} + \beta_5 h_t^{hp} + \beta_6 CCT + \beta_7 Mills + \epsilon_t^w \tag{2.9}$$

[12]Note that those teenagers who are working they do mainly in the informal market, then the exit from the program due to formal income shock is determined by the work decision of their parents

[13]Compulsory high school (Ciclo básico in spanish).

[14]Non compulsory high school (Bachillerato in spanish).

The shocks structure is as follows:

$$\left(\epsilon_t^s, \epsilon_t^w\right) \sim N(\mu, \Sigma) \quad \mu = (\mu_s, \mu_w) \quad \Sigma = \begin{pmatrix} \sigma_s & 0 \\ 0 & \sigma_w \end{pmatrix} \tag{2.10}$$

The Bellman Equations are shown in Equation 2.11 for each choice and depend on the vector of states S_t, which are C_t the accumulated course, the work hours h^w, the home production hours h^{hp}, the gpa, the age, the CCT (control or treatment) and the shocks (ϵ^s and ϵ^w):

$$\begin{aligned} V_t(S_t) &= \max \mathbb{E}\left[\sum_{\tau=t}^{T} \beta^{\tau-t} \sum_k U_t^k d_t^k / S_t\right] \qquad k = \{sn, sw, nn, nw\} \\ S_t &= \left\{C_t, h^w, h^{hp}, gpa, t, T, \epsilon_t' s\right\} \end{aligned} \tag{2.11}$$

$$V_t(S_t) = U_t^k + \beta \mathbb{E}\Big[V_{t+1}(S_{t+1}) | S_t d_t\Big] \tag{2.12}$$

The value function $t < 18$ of the different choices are:

$$V_t(S_t) = \max\left[V_t^{sn}(S_t), V_t^{sw}(S_t), V_t^{nn}(S_t), V_t^{nw}(S_t)\right] \tag{2.13}$$

As in Attanasio et al. (2010), the value function at $t = 18$ is $V_{18}(S_{18})$ which depends on the educational achievement, given that CS is the completed compulsory school and B is the completed non compulsory school. The parameters are estimated in the model.

$$V_{18}(S_{18}) = \frac{\alpha_1}{1 + e^{-\alpha_2 CS_{18} - \alpha_3 B_{18} - \alpha_4 E_{t-1}}} \tag{2.14}$$

2.4 Estimation

The individuals used to estimate the model can enter in two moments, at the beginning of PANES (then they are in that program for 2 years, and then receive AFAM for a maximum of 4 years) or at the beginning of AFAM (and they are in that program for a maximum of 6 years). They can enter at any age between 12 and 17, but the exit is always at 18 years-old. At the entrance the age distribution, the educational level and treated and control status by program are shown in Tables D.1, D.2 and D.3, and these

characteristics are the initial heterogeneity in the model.

Comparing the background of the teenagers when they enter in the programs, there are some slight differences. The control teenagers have a slightly better educational background than the treated ones. Moreover, the AFAM individuals have a better background than the PANES ones, note that the AFAM program is more extensive than the PANES and, consequently, the AFAM teenagers have a better socioeconomic situation and a better educational background. We can observe in Tables D.2 and D.3, that those teenagers who have not completed primary school are at least 10 points lower when we consider the AFAM population in comparison with the PANES population.

The estimation strategy has two steps. In the first one we estimate out of the model the wage function, the GPA function and transition. The second step is the estimation of a group of parameters within the model through the Simulated Method of Moments (SMM).

The parameters estimated out of the model are shown in Tables D.5 - D.7. In the wage equation (Table D.5) we observe that wages and education are negatively correlated, because the wages are determined by specific experience and those who have more education lack this experience. Home production has a positive correlation with wage because there is a complementarity between the intensity in the labor market and the amount of tasks that the teenager does at home.

The GPA dynamics are shown in Table D.6. Performance in $t-1$ has a positive impact on t. The probability of increasing GPA in t is similar for age, but people over 14 years do not change their probability in t when obtaining F or B in $t-1$. Neither home production time nor the market work time coefficients are significant.

Finally, we perform a multinomial logit to estimate the transition between states in the model. As is expected, not only is there some stability of states between t and $t-1$, but also there are significant movements between nn and nw (on both sides). The probability of losing the CCT by a formal income shock is estimated using the administrative records and setting in p_1 as 5.08% and 3.97% and p_2 as 4.69% and 5.88% at the ages of 12-14 and 15-17 respectively. .

The second step of the estimation is through the SMM minimizing the distance between the simulated moments from the model and the data, weighing with the inverse of the simulated variance of the moments. To construct the list of moments we take into consideration the treated and control group. The first one is defined as those who receive

the CCT in the year of entrance in the model, and the latter one the others. We construct the mean for the four states sn, sw, nn, and nw by age and for both programs (PANES and AFAM). Additionally, we consider as moments the time spent on home production, the time working in the market and the grades by age, program and for control and treated populations. These parameters are presented in Tables 2.7 and 2.8.

Total Hours	
L_1 (Study)	3536
L_2 (Non Study)	5096
Home production hours	
HP_1	312
HP_2	624
Market Work hours	
W_1	520
W_2	1040

Table 2.6: Calibration: Hours per year (L_2 means 14 hours per day), (L_1 is equal to L_2 minus 25 hours to attend school).

The total number of hours that the teenagers have to spend is 98 hours per week. If they attend school, they spend 25 hours in that activity. Then they can choose to do some home production 0, 6 or 12 hours per week and work in the market 0, 10 or 20 hours per week. The values per year are shown in Table 2.6.

	12 -13	14 -15	16-17
γ_t	0.3082	0.5808	0.7613
	(0.000344)	*(0.00235)*	*(0.00255)*
b^n_{1t}	5.42	90.68	90.68
	(0.112)	*(31.76)*	*(31.76)*
b^n_2	8.99	8.99	8.99
	(0.343)	*(0.343)*	*(0.343)*
b^{hp}_{1t}	262.95	121.69	198.03
	(37.15)	*(6.25)*	*(10.34)*
b^{hp}_2	6.89	6.89	6.89
	(0.228)	*(0.228)*	*(0.228)*

Table 2.7: Estimation: Parameters estimated by SMM.

The parameter estimation shows that the leisure values increase with age and obviously with the number of hours that the teenagers have available after the school decision, as is shown in Figure D.1. The value of home production does not show a monotone behavior

with age (Figure D.2). In Figure D.3 there is the value of studying which increases in the GPA and in the grade achievement.

Parameter	Value	Std Deviation
β	0.9152	*(0.0061)*
School utility		
b_1	16382.4	*(895.57)*
b_2	10450.0	*(369.15)*
Enforcement parameters		
η_{PANES}	0.057	*(0.0012)*
η_{AFAM}	0.1109	*(0.0092)*
Shocks : means and standard deviation		
μ_s	0	Calibrated
μ_w	-0.838	*(0.0021)*
σ_s	411.26	*(64.038)*
σ_w	0.5514	*(0.0015)*
Final utility function values		
α_1	23024	*(2059.66)*
α_2	0.5315	*(0.0012)*
α_3	0.5555	*(0.00092)*
α_4	0.5355	*(0.00083)*

Table 2.8: Estimation: Parameters estimated by SMM.

The value that the individuals have at the age of 18 depends on equation 2.14 and the set of parameters in Table 2.8, where α_1 is the parameter of Compulsory High school achievement, α_2 the Non compulsory achievement and α_3 the parameter of each grade achievement. The values are shown in Figure D.4

2.5 Results

In this section, we present how well the model fits with the main moments from the data. In the set of Tables 2.9 - 2.12 we show how well the model fits with the states (*sn*, *sw*, *nn* and *nw*) in both programs for treated and control populations. The model fits well with the only exception being for control populations (at the age of 16 and 17) where there is an underestimation of *sn* and an overestimation of *nn*. As we analyze the initial heterogeneity, these teenagers have better conditions than the treated ones. In the model, using the same parametrization could induce these mismatches. These mismatches are observed in depth in the case of AFAM.

	sn		*sw*		*nn*		*nw*	
	Data	Model	Data	Model	Data	Model	Data	Model
12	1	0.9819	-.-	-.-	0	0.0181	-.-	-.-
13	0.965	0.9930	-.-	-.-	0.035	0.0070	-.-	-.-
14	0.759	0.7216	0.0560	0.0067	0.167	0.1998	0.018	0.0719
15	0.764	0.7703	0.0430	0.0071	0.134	0.0990	0.059	0.1237
16	0.627	0.5929	0.0630	0.0007	0.233	0.2340	0.077	0.1724
17	0.521	0.5470	0.0660	0.0000	0.266	0.2140	0.147	0.2391

Table 2.9: Model: PANES Treated.

In Tables 2.9 - 2.12 we can also observe also that the working condition is a bit overestimated, in particular the teenagers in the model have a lot of incentives to only work, and at the same time the state study and working cannot match the data at that age. In sum, the model can capture well the decision to work or not, but it has more problems to disentangle those who work and study, from those who do not.

	sn		*sw*		*nn*		*nw*	
	Data	Model	Data	Model	Data	Model	Data	Model
12	0.951	0.9978	-.-	-.-	0.049	0.0022	-.-	-.-
13	0.946	0.9601	-.-	-.-	0.054	0.0399	-.-	-.-
14	0.861	0.7271	0.031	0.0077	0.108	0.1967	0	0.0685
15	0.748	0.6994	0.038	0.0202	0.206	0.1233	0.008	0.1571
16	0.630	0.5090	0.076	0.0022	0.219	0.2698	0.075	0.2189
17	0.524	0.5154	0.04	0.0000	0.301	0.2125	0.135	0.2721

Table 2.10: Model fit: PANES Control.

In the comparison of treated and control teenagers, the model captures a slightly higher attendance rate for the treated ones. The construction of the paths depends on the initial heterogeneity, which is better for the control ones. The rewards are also a little higher than the control ones because in the data there is a selection process and those teenagers that attend are the better ones.

The model also properly fits the home production, the work behavior and the grade achievement as is shown in Tables D.8-D.10. In the case of the home production the model can disentangle the teenagers that do less and more than 10 hours per week. The model is able to replicate the trend by age in the two time brackets, in the first one there is no clear trend and in the second one it is clearly increasing by age. In the case of work hours, the model has more problems to fit the market work hours properly, because as we mentioned

	sn		*sw*		*nn*		*nw*	
	Data	Model	Data	Model	Data	Model	Data	Model
12	0.985	0.9917	-.-	-.-	0.015	0.0083	-.-	-.-
13	0.924	0.9936	-.-	-.-	0.076	0.0064	-.-	-.-
14	0.880	0.7742	0.006	0.0058	0.092	0.1565	0.020	0.0635
15	0.805	0.8282	0.025	0.0048	0.131	0.0722	0.039	0.0948
16	0.716	0.6617	0.043	0.0005	0.167	0.1751	0.074	0.1626
17	0.608	0.6114	0.057	0.0000	0.223	0.1553	0.112	0.2333

Table 2.11: Model fit: AFAM treated.

before the model overestimates the more intensive workers, and underestimates the less intensive ones.

	sn		*sw*		*nn*		*nw*	
	Data	Model	Data	Model	Data	Model	Data	Model
12	1	0.9930	-.-	-.-	0	0.0070	-.-	-.-
13	1	0.9717	-.-	-.-	0	0.0283	-.-	-.-
14	0.908	0.7903	0.042	0.0078	0.050	0.1695	0	0.0325
15	0.871	0.7858	0.008	0.0291	0.081	0.1091	0.040	0.0761
16	0.807	0.6252	0.038	0.0031	0.115	0.2604	0.038	0.1113
17	0.767	0.6674	0.048	0.0000	0.146	0.1784	0.039	0.1542

Table 2.12: Model fit: AFAM control.

The model captures well the GPA behavior along the ages of the teenagers as is shown in Table D.10. Not only is the trend well captured where A and B are decreasing, and F is increasing, but also the level. Note that, at the age of 12 the percentages are 14.2%, 54.0% and 31.8% for the GPA, F, B and A respectively and those grades at the age of 17 are 57.0%, 31.6% and 11.3%.

2.6 Policy experiments

In the section we perform policy experiments with the enforcement parameters, firstly we set the enforcement parameters 50 points higher, and secondly we set the enforcement parameter at the maximum. Finally, we perform an experiment reducing the amount of the transfers.

Firstly, we set the enforcement parameters η_{PANES} at 0.557 instead at 0.057, and η_{AFAM} at 0.611 instead at 0.111. In Table 2.13 we observe the effect of the policy exper-

PANES: η_{PANES}: 0.557				
	sn	*sw*	*nn*	*nw*
14	0.0890	0.0018	-0.0345	-0.0562
15	0.0467	0.0015	-0.0520	0.0038
16	0.0480	0.0021	-0.0213	-0.0288
17	0.0338	0.0000	-0.0413	-0.0075
AFAM: η_{AFAM}: 0.611				
	sn	*sw*	*nn*	*nw*
14	0.0671	-0.0048	-0.0280	-0.0438
15	0.0394	0.0123	-0.0363	-0.0154
16	0.0533	0.0029	-0.0297	-0.0265
17	0.0517	0.000	-0.0464	-0.0053

Table 2.13: Policy experiment: Differences between treated and control and policy of increasing the enforcement parameter.

iment in both programs, observing the difference-difference between treated and control, and between policy and benchmark. There is a rise in the teenagers that only study and those who study and work and a fall those who do not study or work and those who do not study and work. These rises are higher in PANES than AFAM for the younger ones, due to the fact that the teenagers in the first program have a better background than in the second one.

Secondly, we set the enforcement parameters at the maximum (η_{PANES} and η_{AFAM}) at 1. In Table 2.14 the effect of the states is shown. The effect is quite big, in the early ages the effect is slightly bigger than in the latter ones. Those teenagers that study more, mainly come from those who neither study nor work in contrast to the last case. Here again the effect is higher in PANES than in AFAM, although in this case it is for ages.

The difference of the policy experiment in home production and work hours[15] can be observed in the first panel of Table 2.15. The most important change is the decrease of home production in the younger ages (12-13 years old). In those ages there are no significant changes among states, because more than 90% of those teenagers attend school anyway, although there are changes in the use of their time. For those over 14, there is an effect of reducing the number of hours in home production, with a fall of those who do 12 hours per week, and a slight increment for the less intensive ones.

In the second panel of Table 2.15, we show the effect over the market work hours,

[15]Similar result are found in the case that η is at 0.557 and 0.6111.

	PANES: η_{PANES}: 1.00			
	sn	*sw*	*nn*	*nw*
14	0.1611	0.0143	-0.166	-0.031
15	0.0987	0.0136	-0.100	-0.033
16	0.0986	-0.0001	-0.120	-0.025
17	0.0969	0.000	-0.144	0.007
	AFAM: η_{AFAM}: 1.00			
	sn	*sw*	*nn*	*nw*
14	0.1218	0.0169	-0.0994	-0.0393
15	0.0716	0.0152	-0.0368	-0.0499
16	0.0912	0.000	-0.0670	-0.0243
17	0.0878	0.000	-0.0687	-0.0191

Table 2.14: Policy experiment: Differences between treated and control and policy of increasing the enforcement parameter at 1.

there is a fall in the share of teenagers that work more (more than 15 hours), and an slight increase of less intensive workers. As we analyzed before, the effect over those who study and work contemporaneously is little, but positive.

	PANES		AFAM	
	0-10 hours	>10 hours	0-10 hours	>10 hours
12-13	-0.1500	0.0003	-0.0922	0.0134
14-15	0.0098	-0.0281	0.0403	-0.0121
16-17	0.0157	-0.0048	0.0074	-0.0063
	PANES		AFAM	
	0-15 hours	>15 hours	0-15 hours	>15 hours
14-15	0.0096	-0.0158	0.0108	-0.0186
16-17	0.0061	-0.0199	0.0091	-0.0273

Table 2.15: Policy experiment: Differences between treated and control and policy of increasing the enforcement parameter at 1.

Regarding, the difference in GPA for the effect of the policy, there are less individuals that fail the course and there is an increment of both A and B for all ages. This positive impact in the GPA is generated by the fact that in the ordered probit that generate the grades, being treated has a positive effect on the positive grades (A and B).

Finally, we perform a policy experiment that decreases the amount of transfer by 50%, The difference between treated and control because of this policy is the decrease of those teenagers that only study in both programs, and there is a split between those who study

	GPA		
	F	B	A
12	0.0074	-0.0057	-0.0017
13	-0.0182	0.0258	-0.0076
14	-0.0893	0.0652	0.0241
15	-0.0674	0.0417	0.0257
16	-0.0463	0.0271	0.0192
17	-0.0496	0.0369	0.0127

Table 2.16: Policy experiment: Differences between treated and control and policy of increasing the enforcement rate at 1.

and work, those who neither study nor work and those who only work. In all these states there is an increment of participation. Then, when the transfers is reduced the teenagers go to the other states, but do not necessary drop-out from formal education.

	PANES			
	sn	*sw*	*nn*	*nw*
14	-0.0344	0.0051	0.0252	0.0041
15	-0.0746	0.0136	0.0218	0.0392
16	-0.0321	0.011	0.0104	0.0205
17	-0.0161	0.000	0.0262	-0.0101
	AFAM			
	sn	*sw*	*nn*	*nw*
14	-0.0503	0.0084	0.0371	0.0047
15	-0.0468	0.0209	0.0141	0.0119
16	-0.0174	0.0026	0.0052	0.0097
17	-0.0055	0.000	0.0073	-0.0018

Table 2.17: Policy experiment: Differences between treated and control and policy of decreasing the transfer by 50 points.

2.7 Concluding remarks

In this paper we develop a dynamic model of the teenagers' use of time, when they are elegible to receive a CCT program. We model not only school attendance, but also the work and the home production behavior. Moreover, we model the enforcement of the CCT program, pointing out the estimation of a parameter that shows how the beneficiaries perceive the enforcement level. We exploit a wide, rich and novel data set combining

administrative records and survey data for Uruguay for two CCT programs, in which one of the conditions was school attendance.

In the model, the decision is taken by the teenagers and their parents, given the utility that each decision brings to the household, but the weight of the teenagers in the decision changes over time. This model captures well the state distribution of the individuals not only among the school attendance and market working conditions, but also the time that they spend on home production. One of the features that we exploit is the large percentage of teenagers that neither study nor work, and this is a challenge at the moment to model.

We perform three policy experiments to assess the role of the enforcement parameter and the amount of the transfer in the household decision. When the enforcement is higher the treated teenagers attend formal education more, especially in the middle ages (14-17), before there is no room to improve given the higher rates. Those teenagers in the early ages cannot attend more but they change their use of time, spending fewer hours on home production. Those teenagers who are legally able to work do it less and with less intensity.

Finally, we perform a third policy, reducing the amount of the transfer by 50%. In this case, the effect in PANES treated is higher than in AFAM and there is only a decrease in those who only study, and an increase in the share of all the other states.

2.8 Appendix

	PANES			AFAM		
Age	Total	Treat	Control	Total	Treat	Control
12	17.4	11.9	5.5	17.0	12.9	4.1
13	16.6	12.6	4.1	15.2	11.9	3.3
14	17.1	11.2	5.9	18.3	13.4	4.9
15	16.1	11.2	4.9	16.6	11.7	4.9
16	17.6	13.6	3.9	16.6	12.3	4.3
17	15.2	11.1	4.2	16.3	12.0	4.3

Table D.1: Age and treatment distribution at the entrance moment.

Age	Primary		High-school					
	Incompleted	Completed	1	2	3	4	5	6
			Total					
12	76.6	23.4						
13	49.8	28.1	22.1					
14	34.5	20.9	26.3	18.3				
15	31.7	14.5	21.0	19.9	12.9			
16	31.1	9.7	18.3	16.6	16.0	8.3		
17	33.6	6.7	15.7	15.3	14.3	9.4	5.0	
			Treat					
12	79.4	20.6						
13	54.3	26.3	19.4					
14	39.0	20.9	24.3	15.8				
15	36.8	14.8	20.5	17.5	10.4			
16	35.8	10.0	18.5	15.7	13.6	6.4		
17	38.2	6.6	16.0	14.5	12.9	8.0	3.8	
			Control					
12	69.0	31.0						
13	38.1	32.8	29.1					
14	23.9	20.8	30.8	24.5				
15	19.7	13.9	22.2	25.4	18.8			
16	20.4	9.0	17.8	18.5	21.4	12.9		
17	23.8	6.9	15.2	16.8	17.2	12.5	7.6	

Table D.2: Education background at the moment of entering in PANES.

Age	Primary		High-school					
	Incompleted	Completed	1	2	3	4	5	6
			Total					
12	67.3	32.7						
13	34.2	21.4	44.4					
14	24.8	9.5	30.1	35.6				
15	18.9	10.5	27.4	22.1	21.1			
16	18.0	9.5	33.3	23.5	13.9	1.8		
17	19.7	4.0	24.0	32.1	17.2	1.8	1.2	
			Treat					
12	69.8	30.2						
13	37.1	18.8	44.1					
14	28.0	10.6	29.6	31.8				
15	19.6	9.6	28.8	19.9	22.1			
16	20.1	9.8	31.7	23.6	13.0	1.8		
17	23.4	4.7	23.4	30.4	16.4	1.3	0.4	
			Control					
12	59.4	40.6						
13	24.2	30.3	45.5					
14	15.8	6.3	31.6	46.3				
15	17.3	12.6	24.4	26.8	18.9			
16	12.5	8.6	37.5	23.1	16.4	1.9		
17	8.2	2.1	25.8	37.1	19.6	3.1	4.1	

Table D.3: Education background at the moment of entering in AFAM.

	All	Under-19	Under-19 (CCT)*
Age (reference:12-13)			
14-15	10.308***	9.148***	6.814***
	[1.690]	[0.900]	[2.445]
16-17	10.032***	9.013***	9.584***
	[1.716]	[0.938]	[2.533]
18-19	15.080***	10.804***	14.615***
	[1.828]	[1.081]	[3.290]
20 or more	19.031***		
	[1.531]		
Sex (1=Male)	-21.199***	-5.332***	-6.896***
	[0.503]	[0.637]	[1.817]
Region (1=Capital city)	-1.300**	1.399*	3952
	[0.525]	[0.781]	[2.696]
Employee (1=Yes)	-5.613***	-2.572**	-6.432*
	[0.540]	[1.066]	[3.560]
Attendance (1=Yes. 0=No)	-5.550***	-3.634***	-5.395**
	[0.962]	[0.876]	[2.708]
Offspring (1=Yes)	12.995***	45.903***	36.838***
	[0.634]	[1.910]	[5.805]
Household Income/100	-0.001	-0.001	0.007
	[0.002]	[0.003]	[0.043]
Constant	16.145***	6.276***	8.971**
	[1.521]	[1.136]	[3.691]
N	9387	1481	196
R-square	0.3061	0.4541	0.4465

* Those who applied to the CCT programs (treated and control)

Table D.4: Home production: OLS.

Dependent variable: wage (15-18 years)	OLS	Heckman	Selection eq. Pr(work=1)
	(1)	(2)	(3)
Age	0.063	0.155	0.224***
	[0.105]	[0.195]	[0.038]
Attendance (1=Yes)	-0.030	-0.339	-0.719***
	[0.262]	[0.612]	[0.092]
Education (Ref: Primary)			
High School (compulsory)	-0.460*	-0.448*	0.022
	[0.245]	[0.245]	[0.099]
High School (not compulsory)	-0.856**	-0.918**	-0.153
	[0.378]	[0.392]	[0.136]
Home Production (Ref: HP=0)			
0-10	0.742**	0.719*	-0.062
	[0.379]	[0.379]	[0.133]
> 10	0.813**	0.740**	-0.122
	[0.336]	[0.358]	[0.120]
Treat (1=Yes)	0.272	0.271	0.013
	[0.220]	[0.219]	[0.082]
2nd Wave (1=Yes)	0.337	0.336	-0.012
	[0.207]	[0.206]	[0.077]
Offspring (1=Yes)			-0.713***
			[0.181]
Constant	1.465	-0.640	-4.094***
	[1.720]	[4.149]	[0.636]
Mills			0.553
			[0.992]
N	302		1568
R-sq	0.064		

Table D.5: Wage equation.

Dependent variable: GPA	Age					
	12-18		12-14		15-18	
	(1)	(2)	(3)	(4)	(5)	(6)
GPA t-1 (ref: F)						
B	0.645***	0.631***	0.883***	0.866***	0.069	0.058
	[0.101]	[0.102]	[0.122]	[0.124]	[0.201]	[0.203]
A	1.720***	1.685***	1.892***	1.863***	1.784***	1.738***
	[0.183]	[0.184]	[0.209]	[0.209]	[0.497]	[0.497]
Grade (ref:1-2)						
3-4	0.192	0.212	0.362**	0.385**	0.018	0.019
	[0.138]	[0.139]	[0.184]	[0.185]	[0.222]	[0.225]
5-6	-0.384	-0.373			-1.063***	-1.045***
	[0.3220]	[0.321]			[0.357]	[0.359]
Sex (1=Male)		-0.219		-0.192		-0.140
		[0.134]		[0.186]		[0.206]
Age	-0.127**	-0.128**	-0.257***	-0.259***	0.169	0.149
	[0.056]	[0.057]	[0.091]	[0.094]	[0.114]	[0.114]
Home Production (Ref: HP=0)						
0-10	0.097	-0.047	0.121	-0.008	0.032	-0.029
	[0.102]	[0.143]	[0.114]	[0.187]	[0.313]	[0.321]
> 10	0.077	-0.009	0.084	-0.119	0.166	0.104
	[0.162]	[0.185]	[0.328]	[0.386]	[0.276]	[0.286]
Region (1=Capital city)		-0.197		-0.296		-0.013
		[0.149]		[0.207]		[0.201]
Treat (1=Yes)	0.119	0.130	0.194	0.206	0.046	0.051
	[0.107]	[0.107]	[0.127]	[0.128]	[0.199]	[0.204]
2nd Wave (1=Yes)		-0.156*		-0.187*		-0.114
		[0.095]		[0.128]		[0.184]
Hours Worked (Ref: HW=0)						
0-15					0.464	0.527
					[0.673]	[0.684]
> 15					-0.208	-0.133
					[0.347]	[0.384]
ν_1	-1.827	-2.117	-4.015	-3.682	2.508	2.021
	[0.736]	[0.744]	[1.273]	[1.208]	[1.866]	[1.870]
ν_2	-0.203	-0.479	-2.21	-1.868	3.866	3.384
	[0.732]	[0.741]	[1.266]	[1.204]	[1.873]	[1.878]
N	623	623	454	454	169	169
Pseudo R-sq	0.147	0.153	0.166	0.174	0.091	0.094

Table D.6: GPA: ordered probit.

Multinomial Logit. 15-18 years old.

	sn base outcome								
	sw			nn			nw		
	(1a)	(1b)	(1c)	(2a)	(2b)	(2c)	(3a)	(3b)	(3c)
State									
sw	0.190**	1.217**	1.116**	-0.552	-0.405	-0.488	1.016	1.166	0.870
	[0.561]	[0.556]	[0.564]	[0.767]	[0.721]	[0.718]	[0.683]	[0.786]	[0.851]
nn	0.482	0.593	0.624	2.160***	2.341***	2.385***	2.567***	2.670***	2.658***
	[0.584]	[0.593]	[0.595]	[0.290]	[0.336]	[0.345]	[0.391]	[0.439]	[0.459]
nw	0.350	0.438	0.456	1.243***	1.202***	1.160***	2.261***	2.176***	1.930***
	[0.383]	[0.391]	[0.394]	[0.227]	[0.242]	[0.245]	[0.267]	[0.337]	[0.348]
Sex (1=Male)			0.321			0.670***			2.251***
			[0.345]			[0.251]			[0.366]
Age		0.153	0.137		0.505***	0.551***		0.878***	1.010***
		[0.164]	[0.169]		[0.097]	[0.103]		[0.143]	[0.153]
Home Production (HP=0)									
0-10		0.611	0.781		-0.328	-0.216		-1.444**	-1.407**
		[0.510]	[0.561]		[0.426]	[0.438]		[0.571]	[0.653]
> 10		0.286	0.496		1.326***	1.639***		0.449	1.260***
		[0.506]	[0.548]		[0.340]	[0.347]		[0.408]	[0.410]
Region (1=Capital city)			-0.503			-0.117			-0.483***
			[0.524]			[0.282]			[0.373]
Treat (1=Yes)		0.099	-0.043		-0.038	-0.150		-0.081	-0.214
		[0.370]	[0.363]		[0.225]	[0.231]		[0.276]	[0.300]
Constant	-2.353***	-5.337**	-5.286*	-1.526***	-10.676***	-11.875***	-2.690***	-17.300***	-21.192***
	[0.228]	[2.720]	[2.743]	[0.159]	[1.588]	[1.732]	[0.267]	[2.329]	[2.587]
N	659	653	653	659	653	653	659	653	653
Pseudo R-sq	0.089	0.175	0.211	0.089	0.175	0.211	0.089	0.175	0.211

Table D.7: Transitions: Multinomial logit 15-18 years old.

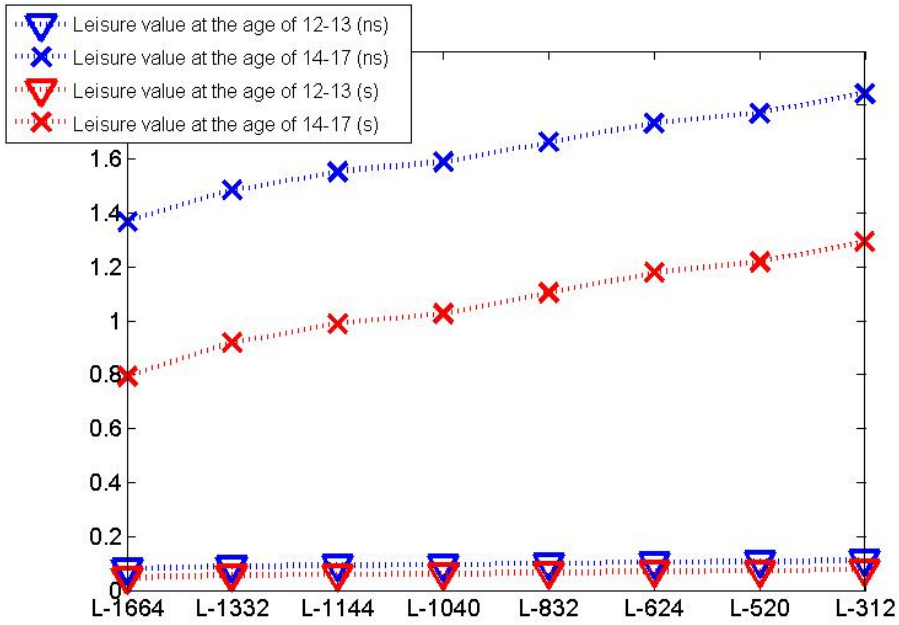


Figure D.1: Leisure value by hours of leisure.

	PANES treated				PANES control			
	0-10 hours		>10 hours		0-10 hours		>10 hours	
	Data	Model	Data	Model	Data	Model	Data	Model
12-13	0.510	0.402	0.054	0.172	0.431	0.362	0.137	0.189
14-15	0.620	0.334	0.358	0.327	0.568	0.347	0.417	0.332
16-17	0.481	0.299	0.390	0.410	0.489	0.326	0.420	0.450
	AFAM treated				AFAM control			
	0-10 hours		>10 hours		0-10 hours		>10 hours	
	Data	Model	Data	Model	Data	Model	Data	Model
12-13	0.507	0.412	0.075	0.156	0.581	0.401	0.071	0.165
14-15	0.599	0.354	0.316	0.311	0.539	0.365	0.395	0.333
16-17	0.531	0.312	0.466	0.408	0.495	0.334	0.505	0.458

Table D.8: Model fit: home production.

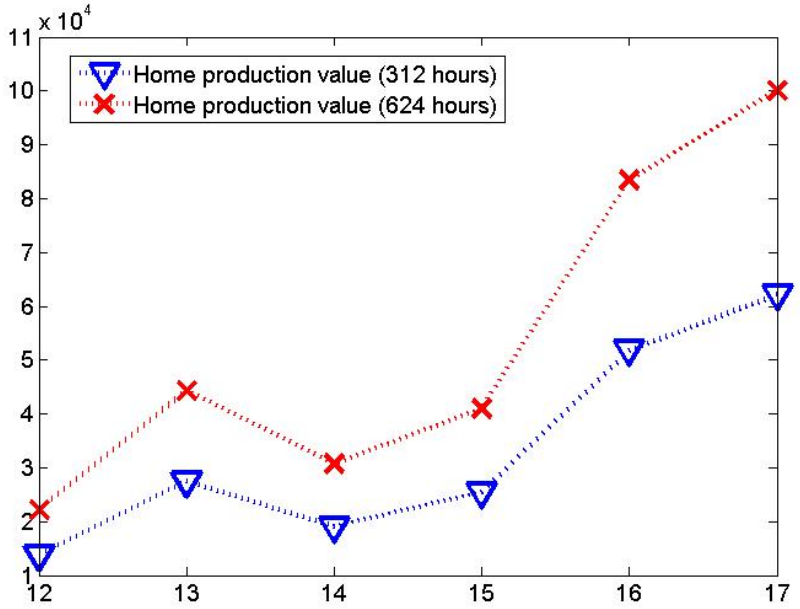


Figure D.2: Home production value by age.

	PANES treated				PANES control			
	0-15 hours		>15 hours		0-15 hours		>15 hours	
	Data	Model	Data	Model	Data	Model	Data	Model
14-15	0.051	0.096	0.036	0.062	0.015	0.086	0.020	0.042
16-17	0.076	0.032	0.116	0.225	0.070	0.025	0.134	0.152
	AFAM treated				AFAM control			
	0-15 hours		>15 hours		0-15 hours		>15 hours	
	Data	Model	Data	Model	Data	Model	Data	Model
14-15	0.021	0.078	0.021	0.054	0.017	0.078	0.029	0.028
16-17	0.045	0.018	0.149	0.218	0.042	0.018	0.151	0.132

Table D.9: Model fit: Work hours.

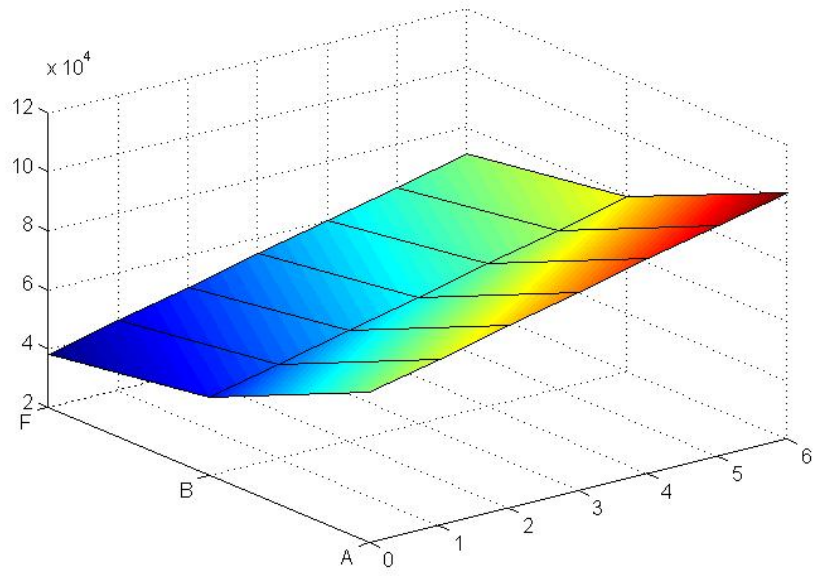


Figure D.3: School attendance value by GPA and grade achievement.

	F		B		A	
	Data	Model	Data	Model	Data	Model
12	0.15	0.145	0.44	0.513	0.41	0.342
13	0.25	0.229	0.52	0.499	0.23	0.272
14	0.31	0.418	0.48	0.399	0.21	0.184
15	0.39	0.393	0.40	0.419	0.21	0.188
16	0.55	0.546	0.34	0.342	0.11	0.112
17	0.57	0.587	0.30	0.306	0.13	0.106

Table D.10: Model fit: Grades by age.

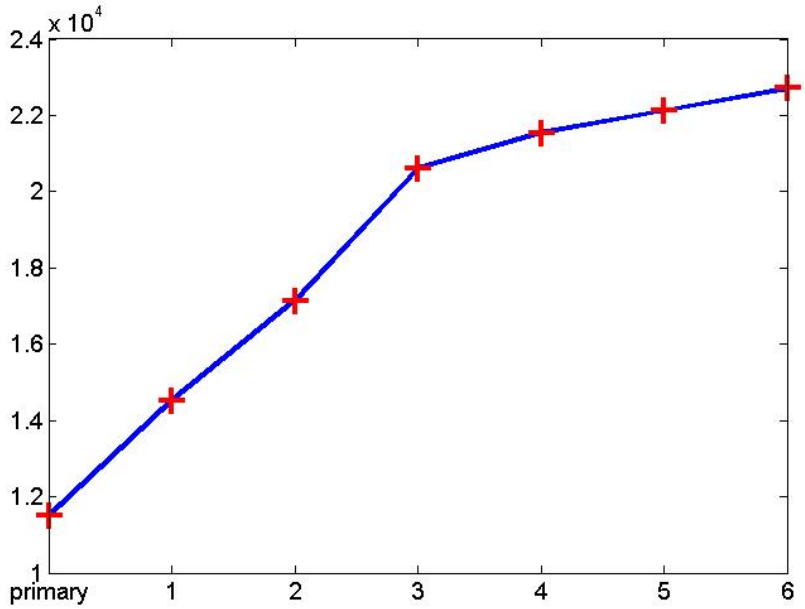


Figure D.4: Value at the age 18.

Chapter 3

Social security schemes and labor supply in the formal and informal sectors

3.1 Introduction

Informality defined as the lack of social security contributions is one of the main characteristics of the labor markets in developing countries. This feature not only has an impact on the current situation of the workers not receiving benefits such as health insurance, unemployment insurance or extra payments, but it also affects access to the pension system for the elderly. This paper discusses **the impact of retirement scheme changes on the labor path between informality and formality and the pension achievement. I will explore the changes in the main variables of the current system, that is: i) number of years of contributing to the system; and ii) the minimum age, which are both requirements to obtain a pension**. Furthermore, this paper deals with the reform of the pension system, from a mixed system where two pillars coexist: a pay-as-you-go (PAYG) system and an individual capitalization system to a new system with only one of those pillars. In order to summarize the different systems I take the rate of replacement as the main characteristic of each scheme, and the different types of pensions (full, advanced age or survival) into consideration.

Over the last few decades, in many developing countries several reforms in the pension scheme have been implemented, aiming to cover the increasing deficit of the public budget triggered by the advance of demographic transition. In South America, in 1975, there were 12.8 working age people (15-64) per each old individual (65 or more); this number fell to 11.4 in 2000 and the estimation for the few next years is an important fall, estimated to be 6.7 for 2025 and 3.7 for 2050 (United-Nations (1999)). This pattern is caused not only by the rise in the life expectancy and survival rates, but also by the fall in birth rates.

Argentina is a special and interesting case because, on the one hand, it is in an advanced stage of demographic transition reaching the same levels as developed countries[1] and, on the other, the system has been the subject of several reforms in the last 20 years. The reform which was established in 1993, transformed a public PAYG with persistent and increasing deficits into a mixed system (PAYG and individual capitalization) in the retirement program in which private and public institutions coexist. This reform was triggered by the need to make the system sustainable. This reform and its consequences were studied in depth by the academia and discussed in the political environments during the last decade, and finally it was reformed in December 2008 to return only to a publicly funded PAYG system[2]. Despite this reform, the requirements to access a pension are still relatively strict in comparison with the region. However, the government carried out a Moratorium in an attempt to drive the universal pension's coverage.

This *Moratorium* was introduced in 2007 as part of the *Pension Inclusion Program*[3] to include mainly women and self employed workers who have eligibility difficulties in the pension system (Bosch and Guajardo (2012), Bosch and Manacorda (2012)). In these papers the *Moratorium* has no effect given that self employed men and women are no included in the estimation, the former because of the lack of measurement through the household survey and the latter because their situation has many ingredients such as fertility and home production which are not included here.

Country	Mean	Male	Female
Uruguay	20.3	13.0	17.7
Chile	23.6	16.6	24.6
Argentina	36.1	27.0	35.5
Colombia	42.7	36.6	38.4
Ecuador	52.0	48.2	45.3
Mexico	55.5	50.5	47.0

Source: CEDLAS

Table 3.1: Informality (lack of pension rights) among salaried workers in LatinAmerica (2009).

Additionally, pension achievement has been in the academic and political discussion in the last two decades. The work record allows the agencies to properly enforce the require-

[1] As is shown in the Figure E.1, Argentina has an advanced stage of demographic transition even in the middle of the past century, and in the projection for 2015 is much closer to the high income countries.

[2] This reform was motivated in part due to the financial crisis after September 2008.

[3] The details can be seen in Arza (2009).

ments to achieve the pension, by checking the years in formality (contribution history). At the same time, those workers who either enter and exit from the formality repeatedly or are in informality for many years have great difficulties in achieving a pension. These workers in the past, even when they did not meet the requirements, were able to easily cheat the agencies through (false) witnesses[4]. Recently in the region[5], there have been some reforms to relax the requirements (reducing the years of contribution, computing, in the case of women, the number of offspring as years of contribution).

The Argentinean program severely punishes short contribution careers in comparison with other countries of the region (Forteza and Ourens (2009)). Conversely, the program also has a wide promotion of extraordinary programs to allow access to specials pensions, for those who do not have enough years to have the right to either the full or the advanced age pensions. Additionally, there has also been an increment in the level of the minimal pension in the last few years (Rofman et al. (2010)). These changes have led to the idea that the pension system is an essential factor in the formality path, because the workers can believe that even if they are in the informality, the government commitment to maintaining some requirement to access a pension will be relaxed (Forteza et al. (2009)). Moreover, if the requirements to access a pension are too strict and only a small share of the population can enjoy it, the government will be forced either to change the rules or to create new types of pensions.

Holzmann and Takayama (2009) focus on the specific effect of some social transfers such as non-contributory pension in the great mobility between formality and informality. Latin America is one of the regions where informality has been studied in depth. About 50% of salaried workers are employed informally, informal workers being defined as those who are not covered by labor regulations, such as taxes, right to the health system and right to pension income in retirement age (Portes et al. (1989), Schneider (2012)). **In this research informal workers are identified as those who declare that their employer is not paying the necessary contribution to have the right to a pension in old age.**

It is relevant to analyze the nature of the informality, in the past the existence of two segmented markets was a common assumption. These markets, formal and informal,

[4]Note the practices are not viewed as a crime in many countries neither by the citizens nor the government.

[5]For example, in Uruguay, Chile or Colombia

Country	Low education	Medium education	High education
Uruguay	27.0	11.0	3.3
Chile	31.8	18.6	12.6
Argentina	51.6	32.0	14.4
Colombia	62.9	30.1	8.2
Ecuador	74.9	47.5	21.6
Mexico	73.9	43,6	24.6

Source: CEDLAS

Table 3.2: Informality Rates Among Salaried Workers by schooling level in 2009.

have different rules which were associated with a high and a low productivity sector respectively. This concept has been discussed by the empirical literature using data from Mexico, Colombia, Argentina and Uruguay[6], where the empirical evidence goes towards the idea that the workers decide where to be, the workers decide which sector to be in based on their characteristics. A third idea proposes a moderate dualism, which is considered in most of the recent theoretical papers[7]. In my paper, I model this moderate dualism through a partial equilibrium model in which the workers decide to work either formally or informally, but being in each sector not only has a different wage function, but also the cost to entry, the cost to change and the probabilities of losing the job are different. This idea is a fundamental concept in this paper.

In the tradition of discrete choice models, Keane and Wolpin (1997) developed a seminal model which provided an estimation of the decision between home production, schooling, and occupational choice. Following this model, Van der Klaauw and Wolpin (2008) developed and estimated a model of labor supply and consumption in low income households with individuals in their fifties. In this model, the individual decides whether to work full or part time or not work at all, subject to the social security rules, limited borrowing, bequests, uncertain health and death. They analyzed single and married individuals separately, and found a lower response among married individuals when social security benefits are reduced by 25%, a reduction in labor supply for individuals below 62 and an increase of total hours for the individual over this age.

Furthermore, Bailey Jones and French (2010) estimate a retirement dynamic model which includes the decision of savings and medical expenses, with special attention paid to the different systems of medical expenses and the role of health insurance. They point

[6]Magnac (1991), Maloney (2004), Pratap and Quintin (2006) and Bucheli and Ceni (2010).

[7]In the Chapter 1 of this Thesis and in Galiani and Weinschelbaum (2007), Amaral and Quintin (2006).

out the relevance of Medicare eligibility in the labor decisions for individuals older than 60 years.

Labor supply behavior in the context of informality has been analyzed in some recent papers, Todd and Velez-Grajales (2008) and Joubert (2012) assess the behavior of individuals among the covered and uncovered sectors[8] for Chile, changing the rules of the pension system. Theoretically, the main differences between these papers and my paper is the definition of informal workers that they are using. They consider as covered (formal) workers those who have a contract, while the uncovered are those who do not have a contract and self employed workers. Additionally, Todd and Velez-Grajales (2008) estimate the model only with men and Joubert (2012) works with couples as a decision unit and he allows for savings in the model. Finally Otero (2013), using data for Chile, analyses the pension system changes in the formality performance with asymmetric intra household bargaining power.

With respect to these last papers, I consider the pension system as a general provision system and not only as a saving system. This allows me to manage different pension schemes and compare them. From an empirical point of view, they are using data for Chile where the rate of uncovered workers is significantly lower than in Argentina. Furthermore, the recent reform in Argentina is an interesting point to study, because the economic volatility gives the role of saving which has a lot of obstacles in the traditional way, to the pension system.

An intertemporal utility maximization model for Brazil using a pseudo-panel is estimated by Robalino (2009). The individual decides where to work and how much to save given a set of social policies. This research has many shortcomings, the main one being the estimation using a pseudo-panel that triggers higher measurement errors. Finally, they do not take data about wages into account nor do they model the longitudinal transition of the workers.

My paper is structured as follows. Section 2 presents some facts from the data to contextualize the topic, Section 3 provides the main features of the Argentinean pension system, Section 4 provides the data which is used in the estimation, Section 5 presents the structural model, Section 6 presents the main results of the estimation, Section 7 the policy experiments, and finally, Section 8 contains the main conclusions.

[8]Covered workers are those who have a written contract. This definition tries to capture some measure of the informality.

3.2 Some facts from the data

Informality is present in all the countries of the region with different degrees depending on the level of development and the institutional framework. In Tables 3.1 and 3.2 the level of informality in six countries in Latin America is shown using the lack of contribution to achieve a pension in the future as the definition[9]. We can identify three groups: Argentina is in the middle one with more than 35% of the salaried workers in informality. These differences are due to the general development level of the countries and the institutional framework such as the quality of the benefits or the level of government enforcement[10]. However, this problem is also present with relevant figures in the most advanced countries of the region: Chile and Uruguay[11]. In the countries with lower informality, there are greater differences between women and men than in countries with higher levels (Table 3.1). Furthermore, differences between educative levels is a feature which is present in all the countries (Table 3.2).

In order to analyze the mobility behavior among formality, informality and unemployment, I perform a multinomial logit with Argentinean data from the period 1995-2008. In Table 3.3 we can observe that the marginal effects, the age, the education and the tenure have a positive effect on formality and a negative one on informality. Being married and being the head of the household has a positive effect of being formal and a negative of being informal, meanwhile being single has a negative effect in both sectors. Being formal in the previous period has a relevant effect on remaining in formality in the current period. It is easier to enter in formality from unemployment than informality relative to being informal in $t-1$, and it is easier to lose an informal job than a formal one.

Based on the multinomial logit model, I predict the distribution in each sector by education, age and marital status. Firstly, the distribution by education is clear, the prediction shows that formality is increasing and informality and unemployment are decreasing by education level, as is shown in Table 3.4. The probability of being unemployed in the lower level of education (incomplete high school) is double in the highest one (college complete). Comparing the highest education with the lowest one, the probability of

[9]Similar figures can be observed when other definitions such as heath insurance or paid holidays are taken into consideration.

[10]These features are treated and discussed deeply in Chapter 1 of this Thesis.

[11]Chile leads almost all the rankings about economic performance and economic development in Latin America.

Marginal effects 1995-2008

	Unemployed	Formal	Informal
Unemployed(-1)	0.0748^{***}	0.0466^{***}	-0.1215^{***}
	(0.0015)	(0.0032)	(0.003)
Formal (-1)	-0.0184^{***}	0.3322^{***}	-0.3138^{***}
	(0.0014)	(0.0010)	(0.0014)
Age	0.0012^{***}	0.0004^{***}	-0.0016^{***}
	(0.0001)	(0.0001)	(0.0001)
Education	-0.0036^{***}	0.0706^{***}	-0.067^{***}
	(0.0011)	(0.0015)	(0.0017)
Married	-0.0095^{***}	0.0273^{***}	-0.0178^{***}
	(0.0035)	(0.0045)	(0.0049)
Single	0.022^{***}	-0.0153^{***}	-0.0068
	(0.0037)	(0.0049)	(0.0056)
Head	-0.0213^{***}	0.043^{***}	-0.0217^{***}
	(0.0020)	(0.0025)	(0.0028)
Tenure	-0.0350^{***}	0.0539^{***}	-0.0189^{***}
	(0.0005)	(0.0008)	(0.0008)

Standard Errors in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 3.3: Marginal effects based on the multinomial model (only men).

being formal is two times higher and the probability of being informal is four times lower. The change in the probabilities when the workers achieve a college degree is remarkable[12]. The probability of being unemployed decreases by more than 4 points, the probability of being formal increases by almost 20 points and being informal decreases by more than 15 points.

The data for age groups shows that the probabilities for the unemployed decreases until the fifties and then increases, the formality has the opposite behavior, increasing and then slightly decreasing. Meanwhile, informality decreases with age but the percentage remains stable after the age of 45. Being formal is 20 points more probable for those over 35 years old than the workers in their twenties, and being informal is 15 points less probable. Regarding the distribution by marital status, the single people have double the probability of being unemployed or 50% higher probability of being informal than those who are married or divorced. In general, the characteristics of formality, informality and unemployment are in line with the literature (Loayza et al. (2009), Hazans (2011))

[12]Medium education means completed high-school and high education completed college.

Distribution in each sector 1995-2008

Education	**Unemployed**	**Formal**	**Informal**
Low Education	0.1137	0.4796	0.4067
Medium Education	0.0861	0.6594	0.2545
High Education	0.0423	0.8578	0.0999
Age group	**Unemployed**	**Formal**	**Informal**
23-28	0.1143	0.5157	0.3700
29-34	0.0767	0.6350	0.2883
35-44	0.0640	0.7048	0.2312
45-54	0.0652	0.7254	0.2094
55-65	0.0829	0.7091	0.2080
Marital status	**Unemployed**	**Formal**	**Informal**
Married	0.0629	0.6949	0.2422
Divorced-widow	0.0850	0.6372	0.2779
Single	0.1371	0.4875	0.3754
Total	**0.0894**	**0.6211**	**0.2895**

Table 3.4: Distribution in each sector by education, age group and marital status based in the multinomial model (only men).

The transitions are studied through the multinomial model which was presented before. Figure E.2 in the Appendix shows the transitions yearly from the formality (first panel), and the informality (second panel). The formality is the sector where the workers remain (do not change) more, but those who change go more to other jobs in the informality rather than to unemployment. In the second panel, 30% of the informal workers annually change sector; in 2003 there is an equal percentage when workers go to formality and to unemployment, and in the latest years the major percentage is for those who change to formality (rather than unemployment).

In Figure E.3, the first panel shows the transition from unemployment and the second panel shows those workers who do not change from the sector who were working the year before (*the stayers*). The unemployed workers tend to change more to informality than to formality and it is easier for the unemployed to enter in the labor market through the informal sector[13]. Formal workers tend to remain in formality more than the informal workers do in informality and the unemployed in unemployment.

Additionally, these transitions are studied by other dimensions. Table 3.5 shows the probability of transitions of the active salaried workers. In the first block, in all the

[13]This feature is also observed if smaller periods are considered, such as quarterly or biannual changes.

Probability of being in each sector			
	Unemployed	Formal	Informal
Unemployed (-1)	0.3797	0.2270	0.3934
Formal (-1)	0.0282	0.8916	0.0802
Informal (-1)	0.1123	0.2140	0.6737
Probability of being unemployed by education and sector of precedence			
Education	Unemployed (-1)	Formal (-1)	Informal (-1)
Low Education	0.3947	0.0417	0.1121
Medium Education	0.3697	0.0266	0.1118
High Education	0.3429	0.0141	0.1154
Probability of being formal by education and sector of precedence			
Education	Unemployed (-1)	Formal (-1)	Informal (-1)
Low Education	0.1758	0.8362	0.1704
Medium Education	0.2531	0.8984	0.2528
High Education	0.3834	0.9492	0.3920
Probability of being informal by education and sector of precedence			
Education	Unemployed (-1)	Formal (-1)	Informal (-1)
Low Education	0.4295	0.1221	0.7175
Medium Education	0.3771	0.0751	0.6354
High Education	0.2737	0.0367	0.4926

Table 3.5: Probabilities of being in each sector, based on the multinomial model 1995-2008 (only men).

elements of the principal diagonal, there are the individuals who do not switch annually[14]. The rows in the table are the original sector where the workers have been in the previous year $(t-1)$, and the columns are the current sector (t). It is important to note that the formality is 20 points more stable than the informality, and about 38% of the unemployed remain in this condition for two consecutive years. At the same time, the informal sector appears more unstable than the formal sector, and it is easier to enter from unemployment (40% instead of 23%), this feature gives some attractiveness to this sector. Additionally, there are 20% of informal workers who annually move to the formality after, for example, gaining some experience or escaping from unemployment.

These changes can also be analyzed by the education level. In the second block, we can observe the transition from unemployment. The stayers in unemployment decrease slightly by educative levels (40%, 37% and 34% respectively); for those who change to the formality the percentages are quite low, but anyway the more educated workers have an

[14]In the whole period 1995-2008.

almost three times lower probability of losing their job (4.2%, 2.6% and 1.4% from low to high educative level). In the case of the informal workers, they can lose their job with the same probability for all educative levels (about 11%).

The third block of Table 3.5, shows the transitions to formality. It is shown that formality is remarkably more stable in the more educated individuals (10 points higher), and the educated workers who were in informality can change to the formality with much more probability than the low educated workers (22 points more). For the unemployed workers, the percentages are 17.6% for the low educated, 25.3% for the medium educated, and 38.3% for the higher ones. Formality has fewer barriers to entry for the high educated than for the other workers.

In the fourth block, the transitions to informality can be observed. Informality is more stable in the lower educative levels, 71.8% for the low educated and 49.3% for the high ones (almost 22 points of difference between the higher and the lower ones). The less educated workers also present more probability to enter to work informally when they used to be unemployed than those with a high educative level (43% and 27% for the lower and the higher educative levels respectively). The same pattern is presented for those who came from formality (12% and 4% for the lower and the higher educative levels respectively).

The distribution of the wages is shown in Figure 3.1 in the Kernel simulations by educative level. The mean of the wages is always higher in formality than in informality in the different levels. However, the informal wages are definitely more volatile (higher standard deviations) than the formal ones, especially in the higher levels of education.

The share of elderly men who achieve some pension payment is shown in Table 3.6. In the data in 2001, most men achieved some pension payment after 75 years old. However, this data is only a photo in 2001 and these workers worked at a time prior to the existence of working records, and the pension requirements were confirmed only by witnesses. If we analyze the work histories Forteza et al. (2009) in the pre-2008 pension scheme in Argentina the simulation shows that only 40% of men would achieve at least 30 years of contributions. Bosch and Guajardo (2012) show that the share of over 65 with a pension income decreases between 1992 and 2007 from about 85% to 65% among men. Principally, those men who have no pension coverage are mainly self-employed.

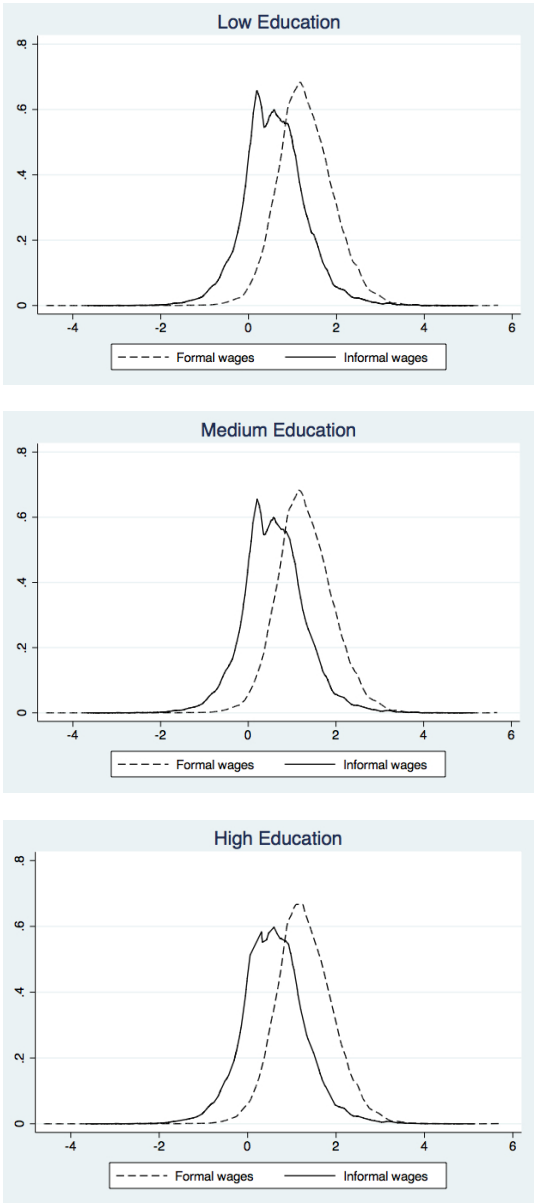


Figure 3.1: Distribution of wages 1995-2008 by education level

Coverage of the pension system

	65-69	70-74	75+
Contributive	64.18	81.8	84.52
Full pension	60.06	75.99	75.37
Advanced age	3.31	4.95	8.16
Non contributive pension	1.31	2.35	3.87

Table 3.6: Coverage of the pension system (only men). Source: Bertranou (2001)

3.3 The Argentinean background: pension system and savings

In Argentina, the pension system has changed drastically twice in the last twenty years. The system changed first in 1994, and a multipillar system based on a PAYG and individual capitalization was established. The first pillar was a PAYG scheme, which was financed by employer's contributions (16% of gross taxable income), and the workers would obtain a Universal Pension Benefit (UPB) with 30 years of contribution and at 60 or 65 years old for women and men respectively. The pension payment was a monthly flat amount of about 28% of the average wage.

The second pillar was financed by employee contributions (11% of gross taxable income), which financed the PAYG or individual capitalization scheme. Private and public institutions participated in this scheme Rofman (2000). There were also employer and employee contributions for different funds to finance redistribution programs and the health system.

In December 2008, as a consequence of the global financial crisis and after more than a decade of criticism of the multipillar regime, the scheme changed again, returning to a single public pillar with a PAYG scheme. The pension is composed of the UPB and Compensatory Pension (CP), which is included to compensate the elimination of individual capitalization. This system is financed by current contributions and general taxes. The employee and employer contributions have not changed, only the administrator of the resources has changed. In the model estimation, I will consider the period 1995-2008 in order to capture the first scheme and the period 2008-2011 as the out of sample validation.

In this paper the different schemes are summarized by the rates of replacement. These figures were estimated taking the different schemes into consideration, and given the mul-

tiplillar scheme was only active for 14 years, there was no possibility to observe an entire generation in the system during its whole active life.

The severe financial crisis that Argentina suffered in 2001 deeply affected saving decisions, because many banks closed and the savers lost a lot of their money. Moreover, in the last two years many restrictions have been imposed to prevent private access to foreign currency which has traditionally been the main way for families to save. However, a pension system is related with the saving decision for the elderly, based on the events of the last decade I have decided not to take into account saving decisions in the model.

3.4 Data

I use the Permanent Household Survey (EPH in Spanish) carried out by the National Institute of Statistics and Census (INDEC in Spanish) for the period 1995 to 2011. The sample is restricted to the urban regions, covering 28 large urban centers where 70% of the urban population of Argentina live[15].

Between 1995 and 2002 the survey was biannual[16], in 2003 it became quarterly. In the first period, the panel is rotative losing 25% of the cases every six months. In the second period, the rotation has the following characteristics: i) two consecutive quarters share 50% of the cases, ii) two quarters with one quarter in the middle, do not have any cases in common, and iii) two quarters with two quarters in the middle, share 25% of cases. Any quarter shares 25% of the cases with the same one in consecutive years. In the whole period it is possible to follow some individuals for one year and a half.

This survey has a socioeconomic purpose and it is crucial in identifying workers in different sectors of the economy. The identification of the formal workers is directly assessed by asking if the employer pays the social contribution to have the right to access a pension payment in retirement. Unfortunately, the questionnaire does not ask anything about the contribution of the self employers. This is the main shortcoming of this survey, so my research only analyzes the dynamic of the salaried workers. This feature allows me to analyze the pure transitions from job to job without taking into consideration self-employment as a possible escape from unemployment.

[15]Urban population accounts for 90% of the total population of Argentina, so the survey gives a good representation of the country.

[16]First and third quarter.

3.5 Model

The model describes the decision problem of the individual in the subsequent periods after they leave the education system until they die. In each period the individuals choose between either working in the formal or informal sector or being unemployed. The individuals have an endowment of human capital which was acquired in the past and depends on the years of schooling, and also the experience that they acquire to work in formal and informal jobs. This latter point means that the employer can not distinguish the sector where the experience of the workers has been gained. For instance, I assume that a young individual at 23 years old leaves the education system with a level of formal education between incomplete elementary school and university degree. They face a finite horizon decision and choose among the different options that they have as in the seminal paper of Keane and Wolpin (1997).

The worker's life is classified in three stages as is shown in Figure 3.2, the first one is the **pure active life**, where the worker can either work in formality or informality, or be unemployed. The second stage, is the **elective retirement** stage where those workers who are eligible for retirement can choose to retire, and those who are not eligible continue as in the pure active life stage. The third and final stage is the **compulsory retirement** stage where everyone is in retirement even if they are not eligible for a pension. In this last stage, everyone receives a pension.

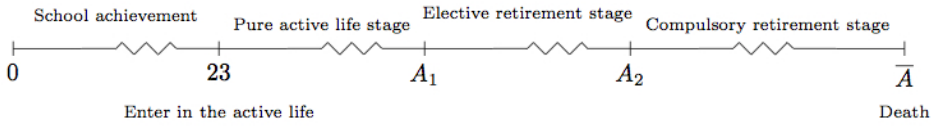


Figure 3.2: Timeline of the individual in the model

Firstly, Equation (3.1) shows the state space, that is: the age (a) which determines the stage of the worker's life; the education level (S) which introduces initial heterogeneity to the individuals, the experience in the labor market (X) which is gained if the individuals are working either in formality or informality, the number of years in formality which is accumulated with the experience in the formal sector (a_F) giving the insight to achieve

one of the requirements to access pension rights, the sector where the individual was in the previous period ($\mathbf{I}_k^{t-1}$) which determines the type of benefit to enjoy (i.e. the unemployment insurance if they choose unemployment after being formal), and the cost of entry to the formal sector (i.e. the cost from unemployment is higher than from informality), and a random shock in wages ($\epsilon(a)^j$).

$$\begin{aligned} \Omega(t) &= \left[a, S, X, a_F, \mathbf{I}_k^{t-1}, \epsilon(a)^j\right] \\ k &= \left\{\text{Formal (F)}, \text{Informal (I)}, \text{Unemployed (U)}\right\} \\ j &= \left\{\text{Formal (F)}, \text{Informal (I)}\right\} \end{aligned} \tag{3.1}$$

There is a Mincer equation of working in each sector in Equation (3.2) where the reward depends on the education (S) and the experience (X) gained working (both in formality and informality), and benefits $B_1(S)$ in formal workers. Both formal and informal wages are hit by an idiosyncratic shock (ϵ^f and ϵ^i).

$$\begin{aligned} \ln\left(R^j(S,X)\right) &= \alpha_0^j \ + \ \alpha_1^j S \ + \ \alpha_2^j X \ + \ \alpha_3^j \left(\frac{X}{10}\right)^2 \ + \ \epsilon(a)^j + B_1(S) \\ j &= \left\{\text{Formal (F)}, \text{Informal (I)}\right\} \qquad \epsilon^j \sim N\left(0, \Sigma\right) \quad \Sigma = \begin{pmatrix} \sigma^f & 0 \\ 0 & \sigma^i \end{pmatrix} \end{aligned} \tag{3.2}$$

The reward of being unemployed is determined by the fact that the worker has the right to enjoy unemployment insurance or not, as is shown in Equation (3.3). If the worker works in formality in t, then the worker can enjoy the insurance in $t+1$, which is fixed as a percentage (b_2) of the previous wage, and they also enjoy $B_2(X,S)$ which reflects leisure or home production.

$$R^U(a,X,S) = \begin{cases} b_2 \mathbf{E}\left[R^F(a-1)\right] + B_2(X,S) & \text{if } \mathbf{I}_F^{-1} = 1 \quad \text{with} \quad 0 \leq b_2 \leq 1 \\ B_2(X,S) \qquad \text{otherwise} \end{cases} \tag{3.3}$$

The functional form of the home production-leisure function increases with education and decreases with experience in the job market as in Equation 3.4. The parameters b_{21}

and b_{23} depend on the educational achievement.

$$B_2(X,S) = b_{21}(S)\left(b_{22} + \frac{b_{23}(S)}{X}\right) \tag{3.4}$$

The **first stage of analysis** is the pure active life until age A_1, the individual decides taking not only the current and the future value function, but also the transition probabilities into account. The transition probabilities are defined as $\lambda_f(S)$, which is the probability of formal job destruction (formality $\rightarrow$ unemployment), $\lambda_i(S)$, which is the probability of informal job destruction (informality $\rightarrow$ unemployment), $\phi^F(S)$ (unemployment $\rightarrow$formal) which is the probability of finding a job in the formal sector, and $\phi^I(S)$ (unemployment $\rightarrow$ informal) which is the probability of finding a job in the informal sector. The value function V^j in Equation (3.5) is the maximum among the value of being employed in the sector j, and V^U which is the value of being unemployed.

Equation (3.6) shows the value function of working in formality V^F (informally V^I) is defined with the formal (informal) Mincer equation, the cost of entering each sector, the cost of switching sectors and the expectation of the future depends on the probability $\big(1-\lambda_k(S)\big)$ of being employed and the future realization of the wages, that is the expected discounted value of all the years of the first stage (pure active life), plus the value function of the second and the third stage. The cost of entering and switching sectors depends on the sector of the previous and the current period (see Equation 3.8).

$$V\big(\Omega(t)\big) = max\Big\{V^F, V^I, V^U\Big\} \tag{3.5}$$

$$\begin{aligned} V^j\Big(\Omega(a)\Big) &= U\Big(R^j(a) - \mathbb{C}^{Fk}(a,S)\mathbf{I}^{-k}\Big) \\ &+ \beta\Bigg[\lambda_j(S)\mathbf{E}_{\Omega(a+1)/\Omega(a)}V^U\Big(\Omega(a+1)\Big) \\ &+ \big(1-\lambda_j(S)\big)\mathbf{E}_{\Omega(a+1)/\Omega(a)}max\Big\{V^I\Big(\Omega(a+1)\Big), V^F\Big(\Omega(a+1)\Big)\Big\}\Bigg] \\ j &= \Big\{\text{Formal (F)}, \text{Informal (I)}\Big\} \end{aligned} \tag{3.6}$$

The cost of entry in formality (Equation 3.7) depends on the educational endow-

ment(S), the age (a), where the worker was in the previous period ($\mathbf{I}^{-k}$ i.e. working informally or unemployed) and a set of parameters $(\Pi, \phi_{11}(S), \phi_{12}(S))$. The cost of entry in informality is fixed at zero.

$$\mathbb{C}^{Fj}(a, S, \mathbf{I}^{-k}) = \begin{cases} (\Pi - a)\phi_{11}(S) & \text{if} \quad \mathbf{I}^{-k} = \text{informal} \\ (\Pi - a)\phi_{12}(S) & \text{if} \quad \mathbf{I}^{-k} = \text{unemployed} \\ 0 & \text{if} \quad \mathbf{I}^{-k} = \text{formal} \end{cases} \tag{3.7}$$

$$k = \big\{\text{Formal (F), Informal (I), Unemployed (U)}\big\}$$

$$\mathbf{I}^{-k} = \begin{cases} 1 & \text{if the individual being in } -k \text{ in the previous period} \\ 0 & \text{otherwise} \end{cases} \tag{3.8}$$

The value function for the unemployed worker depends on the current utility function, which is different if in the previous period the workers were in formality or in informality, and the expected utility function, as is shown in Equation 3.9.

$$\begin{aligned} V^U\Big(\Omega(a)\Big) &= U\Big(R^U(a)\Big) \\ &+ \beta\Bigg[\Big(1 - \phi_F(S) - \phi_I(S)\Big)\mathbf{E}_{\Omega(a+1)/\Omega(a)}V^U\Big(\Omega(a+1)\Big) \\ &+ \phi_F(S)\mathbf{E}_{\Omega(a+1)/\Omega(a)}max\Big\{V^U\Big(\Omega(a+1)\Big), V^F\Big(\Omega(a+1)\Big)\Big\} \\ &+ \phi_I(E)\mathbf{E}_{\Omega(a+1)/\Omega(a)}max\Big\{V^U\Big(\Omega(a+1)\Big), V^I\Big(\Omega(a+1)\Big)\Big\}\Bigg] \end{aligned} \tag{3.9}$$

The utility function takes a general form, Constant Relative Risk Aversion (CRRA) as in Equation 3.10:

$$U\Big(R^k(a)\Big) = \frac{1}{1-\gamma}R^k(a)^{\frac{1}{1-\gamma}} \tag{3.10}$$

The **second stage of analysis** is between A_1 and A_2, A_1 is the lower age that the workers can choose to retire and at A_2 everyone is retired. In this period the individuals can choose to continue to work in both sectors or be unemployed, but those individuals who achieve the minimum years in formality (F_1) may get a full pension, so they could be retired and also enjoy $B_3(S)$ as pensioner's home production-leisure as in Equations

3.11 and 3.13. Otherwise, the pension could be achievable in this period when a_F (total of years worked in formality) achieve at least the threshold F_1. Note, as Equation 3.12 shows the pension is always a choice in this stage of the workers life. Additionally, there are three full pensions which depend on a_F, if the workers decide to continue working, they can achieve a higher rate of replacement in the future, but their life is limited to $\overline{A}$ when everyone is dead.

$$V^P\Big(\Omega(a)\Big) = U\Big(r_F R^k(a) + B_3(S)\Big) + \beta \mathbf{E}_{\Omega(a+1)/\Omega(a)} V\Big(\Omega(a+1)\Big) \tag{3.11}$$

$$V\Big(\Omega(a)\Big) = max\Big\{V^U\Big(\Omega(a)\Big), V^I\Big(\Omega(a)\Big), V^F\Big(\Omega(a)\Big) V^P\Big(\Omega(a)\Big)\Big\} \tag{3.12}$$

The home production leisure function (Equation 3.13) for the pensioners depends only on educational achievement and it is constant during the retirement period.

$$B_3(S) = b_{31}(S) \tag{3.13}$$

The individuals who achieve the full pension are those who work in formality for 30 years or more:

$$\mathbf{I}_{a_F \geq 30} = \begin{cases} 1 & \text{if the individual has worked in formality for 30 years or more} \\ 0 & \text{otherwise} \end{cases} \tag{3.14}$$

The **third stage of analysis** starts at A_2 years old when all the individuals are retired. The value function of these pensioners is shown in Equation 3.15 and is determined by the income that the individual would receive and the number of years in formality (F_1 and F_2). It is determined by the replacement rate and the last wage received in the active life. There would be three types of pensions: the full, the *Advanced age* and the survival pension. At the age of $\overline{A}$ everyone is dead.

$$\begin{aligned} V\Big(\Omega(a)\Big) = {} & U\Big(r_F R^k(a) + B_3(S)\Big)\mathbf{I}_{a_F \geq 30} + U\Big(r_A R^k(a) + B_3(S)\Big)\mathbf{I}_{30 > a_F \geq 10} \\ & + U\Big(b_3 + B_3(S)\Big)\mathbf{I}_{a_F < 10} + \beta \mathbf{E}_{\Omega(a+1)/\Omega(a)} V\Big(\Omega(a+1)\Big) \end{aligned} \tag{3.15}$$

The individuals who achieve the *Advanced age* pension are those who work between

10 and 30 years in formality during their whole working life:

$$\mathbf{I}_{30>a_F\geq 10} = \begin{cases} 1 & \text{if they have worked in formality for 10 years or more and less than 30} \\ 0 & \text{otherwise} \end{cases} \tag{3.16}$$

The rate of replacement (r_F) in the full pension type is not unique, it increases in the number of years that the workers one in formality. In the estimation I will consider three rates, at 30 (a_{F-30}), 35 (a_{F-35}) and 40 (a_{F-40}) years in formality.

3.6 Estimation

I estimate the parameters of model using the Simulated Methods of Moments (SMM), through the maximization of the value function conditional on the state variables minimizing the distance between the simulated moments in the model and the moments in the data, weighing with the inverse of the simulated variance of the moments. I select moments to match the choices along the working career and the educational endowment and the transition between states, to estimate the discount factor β, the shape of the utility function γ, the parameters that determine the home production-leisure function, the parameters to enter into formality from informality and unemployment and the idiosyncratic shocks.

In Table 3.7 I show the estimation of the Mincer equation for the formal and informal workers, each worker based on their educational endowment and the experience that they achieve in the labor market, receives a wage offer which forms part of their utility function and decides either to work or to remain unemployed. Education and experience are better rewarded for the formal workers than for the informal ones. However, the informal workers receive better offers in their youth (the constant term is higher). The standard deviation of the idiosyncratic shocks in the wages is estimated by MSM and is shown in Table 3.8, informal shocks have a higher deviation than formal ones.

The transition function (λ_j and ϕ_j) parameters (education endowment and experience achieved) are estimated through the marginal effect of the multinomial function shown in Table 3.3.

The parameters estimated by SMM are shown in Table 3.8. In the first block are the general parameters of the model, the discount factor which is in line with the literature,

Coefficients estimated from the Mincer equations

Name	Symbol	Value	Standard deviation
Constant formal	α_0^f	0.2606	(*0.0227*)
Constant informal	α_0^i	0.3962	(*0.0805*)
Schooling formal	α_1^f	0.2982	(*0.0027*)
Schooling informal	α_1^i	0.2707	(*0.0056*)
Experience formal	α_2^f	0.1622	(*0.0035*)
Experience informal	α_2^i	0.0979	(*0.0057*)
Experience^2 formal	α_3^f	-0.0117	(*0.0004*)
Experience^2 informal	α_3^i	-0.0077	(*0.0007*)

Table 3.7: Parameters of the Mincer equation.

is close to 0.95 and the risk aversion is 1.28 which is a bit lower than in the literature[17]. The second block shows the parameters of the extra-wage benefits that the formal workers enjoy, which is fixed at zero for those in the lower educative level. In the third block, there are the home production-leisure parameters in active and in the fourth one, the parameters in retirement. The parameters of the home production-leisure are higher for the more educated, reflecting the outside option of the wage. In the fifth block, there are the parameters of the cost to enter into formality, the low educated have a higher cost than the higher educated individuals.

Some parameters are calibrated in Table 3.9. The minimum retirement age is 65 years old as the minimum age to achieve the full pension and 70 years old is the age to achieve the *Advanced age* pension. Additionally, to achieve the full pension the workers have to work at least 30 years in formality and 10 years to get the *Advanced age* pension. The full pension has three levels with three different rates of replacement at 30, 35 and 40 years in formality with 1.13, 0.96 and 0.81 being the rates of replacement respectively. The *Advanced age* pension is at 70% of the full pension with 30 years of formality. Everyone is dead at 81 years old.

In Table 3.10 there is the estimation for all education levels, the model fits well if the interval is considered. The model has more problems at the beginning and at the end in the career, and especially between formal and informal workers, with an overestimation of

[17]For instance, Joubert (2012) find 1.55 for Chile.

Parameters estimated by Simulated Method of Moments

Name	Symbol	Value	Standard Deviation
Discount factor	β	0.9475	(*0.0014*)
Risk aversion	γ	1.2757	(*0.0018*)
Benefits $B_1(S)$			
Medium Education	$B_1(2)$	0.4691	(*0.0004*)
High Education	$B_1(3)$	0.9382	(*0.0002*)
Home production in active life $B_2(XX, S)$			
Constant	b_{22}	100.9925	(*5.589*)
Low Education	$b_{21}(1)$	38.1481	(*0.781*)
Medium Education	$b_{21}(2)$	36.9016	(*1.631*)
High Education	$b_{21}(3)$	49.5058	(*2.396*)
Low Education	$b_{23}(1)$	43.465	(*1.090*)
Medium Education	$b_{23}(2)$	51.5392	(*0.925*)
High Education	$b_{23}(3)$	57.3335	(*2.557*)
Home production in retirement life $B_3(S)$			
Survival pension	b_3	1838.1	(*2.0*)
Low Education	$b_{31}(1)$	1060.395	(*1.87*)
Medium Education	$b_{31}(2)$	3182.1625	(*48.91*)
High Education	$b_{31}(3)$	2113.5710	(*112.4*)
Cost of entering in formality			
Age multiplier	Π	97.6199	(*4.09*)
Informal-Formal (Low Education)	$\phi_{11}(1)$	1.18	(*0.0141*)
Informal-Formal (Medium Education)	$\phi_{11}(2)$	0.591	(*0.0*)
Informal-Formal (High Education)	$\phi_{11}(3)$	0.0004	(*0.0*)
Unemployed-Formal (Low Education)	$\phi_{12}(1)$	45.044	(*1.22*)
Unemployed-Formal (Medium Education)	$\phi_{12}(2)$	4.504	(*0.12*)
Unemployed-Formal (High Education)	$\phi_{12}(3)$	0.015	(*0.0*)
Shocks			
Std Variation informal	σ^i	0.3602	(*0.00*)
Std Variation formal	σ^f	0.2109	(*0.00*)

Table 3.8: Parameters estimated by Simulated Method of Moments.

Parameters calibrated

Name	Symbol	Value
Thresholds		
Minimum full retirement age	A_1	65
Advanced age retirement age	A_2	70
Death	$\overline{A}$	80
Years in formality to achieve full pension	F_1	30
Years in formality to achieve advanced age retirement	F_2	10
Rates of replacement		
Full pension (40 years of formality)	r_{F-40}	1.13
Full pension (35 years of formality)	r_{F-35}	0.96
Full pension (30 years of formality)	r_{F-30}	0.81
Advanced age pension	r_A	0.7*r_{F-30}
Unemployment benefits for former formal workers	b_2	0.6

Table 3.9: Parameters calibrated. Source: Forteza and Ourens (2009), Rofman et al. (2010) .

informality in youth and an underestimation close to retirement.

In Table 3.11 the moment matching of the formal activity is shown. The model matches quite well taking into consideration the interval, especially the high educated and the general estimation. However, in the case of the low educated and the medium educated there is an underestimation (more than 10 points) at the beginning and an overestimation (also 10 points) in the last years of the career.

In Table 3.12 the moment matching is considered comparing the informal activity. In this case the model behavior is the other side of the coin, the estimation is again quite good in the general case and in the medium educated workers, and in this case the high educated group is mostly in the interval, but there is an overestimation of the low educated group. The model matching for the unemployed workers is the residual and it is presented in Table E.3.

In Table 3.13, there is the model matching of the transitions. Here the model behavior is also quite good, because the formal workers tend to stay in this sector more than the informal and the unemployed ones. Informality is an easier sector to enter from unemployment than formality. However, the transition from unemployment to informality is underestimated.

In this benchmark case the initial heterogeneity of the model, only the education en-

	Formal		Informal		Unemployed	
	Data	Model	Data	Model	Data	Model
23-28	0.5311	0.3993	0.3615	0.5112	0.1074	0.0895
	[0.5073, 0.5482]		[0.3400, 0.3841]		[0.0962, 0.1163]	
29-34	0.6286	0.6194	0.3091	0.3505	0.0623	0.0301
	[0.5979, 0.6534]		[0.2768, 0.3415]		[0.0510, 0.0722]	
35-44	0.7418	0.7564	0.2107	0.2256	0.0475	0.0180
	[0.7148, 0.7621]		[0.1958, 0.2295]		[0.0398, 0.0574]	
45-54	0.7104	0.7466	0.2218	0.1947	0.0678	0.0588
	[0.6870, 0.7277]		[0.2143, 0.2299]		[0.0507, 0.0943]	
55+	0.7127	0.8314	0.2067	0.0836	0.0807	0.0849
	[0.6882, 0.7374]		[0.1890, 0.2359]		[0.0603, 0.1100]	

Table 3.10: Moment matching: Formal, informal and unemployed workers at all educational levels. The interval is estimated based on the multinomial model, performing bootstraps and considering the middle 90%.

	Low Education		Medium Education		High Education	
	Data	Model	Data	Model	Data	Model
23-28	0.4048	0.2654	0.6001	0.4811	0.6754	0.6150
	[0.3628, 0.4356]		[0.5863, 0.6154]		[0.6310, 0.7167]	
29-34	0.4788	0.5115	0.7061	0.6917	0.8839	0.7783
	[0.4429, 0.5081]		[0.6682, 0.7324]		[0.8492, 0.9187]	
35-44	0.5959	0.6514	0.8474	0.8515	0.9186	0.8538
	[0.5739, 0.6185]		[0.8018, 0.8833]		[0.9041, 0.9315]	
45-54	0.5761	0.5440	0.8594	0.9432	0.9290	0.9037
	[0.5088, 0.6290]		[0.8119, 0.8955]		[0.9049, 0.9468]	
55+	0.6620	0.7644	0.7081	0.8753	0.9584	0.9327
	[0.6209, 0.6964]		[0.6491, 0.7931]		[0.9364, 0.9779]	

Table 3.11: Moment matching: Formal workers. The interval is estimated based on the multinomial model, performing bootstraps and considering the middle 90%.

	Low Education		Medium Education		High Education	
	Data	Model	Data	Model	Data	Model
23-28	0.4490	0.6485	0.3169	0.4149	0.2438	0.3189
	[0.4045, 0.4932]		[0.3004, 0.3324]		[0.1973, 0.2876]	
29-34	0.4367	0.4523	0.2412	0.2757	0.0988	0.2161
	[0.4066, 0.4743]		[0.2054, 0.2817]		[0.0720, 0.1136]	
35-44	0.3313	0.3145	0.1216	0.1453	0.0688	0.1428
	[0.3142, 0.3486]		[0.0882, 0.1544]		[0.0554, 0.0797]	
45-54	0.3187	0.3469	0.1186	0.0537	0.0549	0.0608
	[0.3003, 0.3379]		[0.0852, 0.1564]		[0.0376, 0.0767]	
55+	0.2518	0.1386	0.1875	0.0322	0.0320	0.0365
	[0.2236, 0.2852]		[0.1468, 0.2383]		[0.0147, 0.0495]	

Table 3.12: Moment matching: Informal workers. The interval is estimated based on the multinomial model, performing bootstraps and considering the middle 90%.

	Unemployment		Formal		Informal	
	Data	Model	Data	Model	Data	Model
Unemployed (-1)	0.4414	0.3228	0.1928	0.2593	0.3658	0.4181
	[0.3668, 0.5152]		[0.1501, 0.2268]		[0.3141, 0.4193]	
Formal (-1)	0.0295	0.0267	0.8824	0.8347	0.0881	0.1388
	[0.0238, 0.0425]		[0.8751, 0.8907]		[0.0761, 0.0974]	
Informal (-1)	0.0743	0.1234	0.2393	0.4755	0.6864	0.4013
	[0.0593, 0.1034]		[0.2142, 0.2673]		[0.6719, 0.7068]	

Table 3.13: Moment matching: Transitions. The interval is estimated based on the multinomial model, performing bootstraps and considering the middle 90%.

At 65 and 70 years old	Full_{40}	Full_{35}	Full_{30}		Advanced age	
			Age 65	Age 70	Age 65	Age70
Low Education	0.0%	1.4%	22.9%	44.4%	75.4%	54.2%
Medium Education	0.3%	40.9%	55.3%	58.5%	3.6%	0.3%
High Education	6.2%	56.4%	35.4%	37.3%	2.0 %	0.0%

Table 3.14: Pension achievement.

dowment and the experience achieved, does not allow for the differentiation of individuals.

Pension achievement is shown in Table 3.14, all the workers achieve at least an *Advanced age pension*, given that the requirements are very loose, only 10 years of contribution or formality. For the low educated workers almost 54% achieve the *Advanced age pension* and 44% the first step of full pension. Note that most of this individuals reach the pension after the age of 65, then they continue working in order to have this right[18]. There is a significant share of medium educated workers who continue to work after 65 in order to achieve a full pension, and only 0.3% get only the *Advanced age* pension. At the age of 70, 58% of these workers have the first step of full pension, and 41% and 0.3% the second and third steps. Most of the high educated workers get the second and third steps of the full pension. In this estimation, the main problem about the pension achievement is in the low part of the distribution. Most of the high educated achieve a pension, meanwhile the low educated have to work until the age of 70 to achieve an *Advanced age* pension. This point will be discussed in the policy experiments.

The distribution of the formality path is shown in Table 3.15. The age to start in formality is lower for higher educated workers, 2 years before, and the average years in formality is almost double when comparing these groups. Most of the workers prefer

[18]At age 65 the only pension that they can achieve is the full pension.

	Low Education		Medium Education		High Education	
	Age 65	Age 70	Age 65	Age 70	Age 65	Age 70
Mean	24.83	27.32	33.98	34.05	35.46	35.51
Start age	4.16		3.13		2.56	
$p_{0.025}$	14	18	29	30	30	30
$p_{0.125}$	18	22	31	31	32	32
$p_{0.25}$	21	25	32	32	33	33
$p_{0.5}$	25	29	34	34	36	36
$p_{0.75}$	29	30	36	36	38	38
$p_{0.875}$	31	31	37	37	39	39
$p_{0.975}$	34	34	39	39	40	40

Table 3.15: Years in formality.

to retire when they have the possibility, it is clear that for high educated workers the distribution is the same at 65 and 70 years old. This is not the case for the medium and low educated workers where the workers about 75% and 25% respectively, continue working after 65. Informality affects the low educated workers more, as they do not enjoy the current and the future benefits.

3.7 Policy Experiments

In this section I want to perform some policy experiments in order to analyze different situations, which could be probable in different scenarios. First I will analyze what situation will be if the rate of replacement was also lower[19] (as in a PAYG scheme) and the age of retirement was lower. Then I make the requirements stricter, and then I perform changes that could affect the distribution, affecting differentially the low and the high educated workers..

The first policy experiment is to reduce the rate of replacement that the workers would get with the full pension[20] and the minimum age to get a pension is 60 years old. In Table 3.16 the formality path is shown, there is a decrement of formality at the end of the career. For the medium educated workers the fall is almost 20 points, and for the low and high educated workers it is 15 points. Those workers decide to be in unemployment more than to work.

[19]In the benchmark calibration I consider the rate of replacement in the case of individual a capitalization scheme, here I consider the rate of replacement of PAYG.

[20]The rates of replacement are now 0.8 and 0.7 in the full pension, the *Advanced age* pension is 70% of the latter one.

	General		Low Education		Medium Education		High Education	
	Benchmark	Policy	Benchmark	Policy	Benchmark	Policy	Benchmark	Policy
23-28	0.399	0.404	0.265	0.274	0.481	0.484	0.615	0.618
29-34	0.619	0.616	0.512	0.503	0.692	0.694	0.778	0.777
35-44	0.756	0.745	0.651	0.649	0.852	0.851	0.854	0.853
45-54	0.747	0.755	0.544	0.556	0.943	0.944	0.904	0.918
55+	0.831	0.677	0.764	0.564	0.875	0.785	0.933	0.770

Table 3.16: Formality path with the first policy experiment: replacement rates and retirement age at 60.

	Full_{35}		Full_{30}		Advanced age	
	Benchmark	Policy	Benchmark	Policy	Benchmark	Policy
Low Education	1.4%	0.0%	44.4%	32.2%	54.2%	67.8%
Medium Education	41.2%	1.3%	58.5%	97.4%	0.3%	1.4%
High Education	62.6%	3.3%	37.3%	96.4%	0.0%	0.3%

Table 3.17: Pension achievement with the first policy experiment: replacement rates and retirement age at 60.

The pension achievement changes are shown in Table 3.17. The main change is the rise from about 60% to 97% in the medium educated who achieve the first step of the full pension. The workers with low education who achieve a full pension is 12 points lower than in the benchmark, this tendency is confirmed in Table 3.18, which shows that the years of formality is lower in the low part of the distribution.

The distribution in Table 3.18 shows that the low educated workers work less formally, especially between the age of 65 and 70. On average, at the age of 70, the low educated workers have about 2 years less in formality and it is explained principally by the low part of the distribution. The medium educated workers work less in formality (almost 3.5 years in average at the age of 70), and it is explained principally by the fact that they prefer to go into retirement as soon as possible, and for this reason the distribution is more homogeneous in this variable with the policy experiment.

The second policy puts stricter requirements to achieve both the full and *Advanced age* pension. The minimum years in formality to achieve them are 35 and 20 respectively. Table 3.19 shows the pension achievement for the three levels of education. In this scenario, there are changes principally again for those with low and medium education, in the former almost 6% do not reach the *Advanced age* pension and only 17% achieve a full pension. For the medium educated, there are changes with more workers (7% instead of 0.3%) that achieve only the *Advanced age* pension and the first step of the full pension, the second

	Low Education			
	Benchmark	Policy	Benchmark	Policy
Years of formality (at 65)	24.8	23.9		
Years of formality (at 70)	27.3	25.6		
Age of start in formality	4.15	4.08		
	Distribution at age 65		Distribution at age 70	
$p_{0.025}$	14	14	18	15
$p_{0.125}$	18	17	22	20
$p_{0.25}$	21	20	25	22
$p_{0.5}$	25	24	29	27
$p_{0.75}$	29	29	30	30
$p_{0.875}$	31	30	31	30
	Medium education			
	Benchmark	Policy	Benchmark	Policy
Years of formality (at 65)	33.9	30.56		
Years of formality (at 70)	34.1	30.74		
Age that start in formality	3.15	3.13		
	Distribution at age 65		Distribution at age 70	
$p_{0.025}$	29	28	30	30
$p_{0.125}$	31	30	31	30
$p_{0.25}$	32	30	32	30
$p_{0.5}$	34	30	34	30
$p_{0.75}$	36	31	36	32
$p_{0.875}$	37	32	37	32
$p_{0.975}$	39	34	39	34

Table 3.18: Formality with the first policy experiment: replacement rates and retirement age at 60.

	$Full_3$		$Full_2$		$Full_1$	
	Benchmark	Policy	Benchmark	Policy	Benchmark	Policy
Low Education	0.0%	0.0%	1.4%	0.0%	44.4%	17.8%
Medium Education	0.3%	0.0%	40.9%	1.2%	58.5%	91.7%
High Education	6.2%	0.0%	56.4%	6.5%	37.3%	91.0%
	Advanced Age		Survival			
	Benchmark	Policy	Benchmark	Policy		
Low Education	54.2%	75.9%	0.0%	6.3%		
Medium Education	0.3%	7.1%	0.0%	0.0%		
High Education	0.0%	2.6%	0.0%	0.0%		

Table 3.19: Pension achievement with the second policy experiment: formality years of requirement 20 and 35. $Full_3$ pension is 43 and 40 years, $Full_2$ pension is 40 and 35 years and $Full_1$ pension is 35 and 30 years in the policy and in the benchmark respectively.

and third steps were reached by about 41% and now this percentage is only 1%.

In Table E.4 there is an increment in the formality among individuals over 55 years old in all educative levels. For the lower and medium educated it is about 7% higher and for the high educated the difference is 1.5%.

The years in formality is shown in Table 3.20. There are changes in the distribution around the new thresholds, among the low educated there is a decrease in the low part of the distribution (18 to 17 in the 2.5 centile) and an increase in the high part of the distribution (30 to 33 in the 75 centile and 32 to 35 in the highest part). Additionally, there are changes for the medium educated, there is an increase in the formality for all the workers of the distribution, but the change is higher in the low part of the distribution where most of the workers work formally in order to achieve the new threshold and enjoy a full pension.

The third policy puts looser requirements to achieve a full pension, only 25 years in formality. This policy decreases the years of formality for all the educative levels especially in the last part of the career (over the age of 45) as in shown in Table E.5. The workers who achieve an *Advanced age* pension is half that of the benchmark even if they work less in formality, because this decrement is after 65 then they reach the new minimum threshold and go into retirement.

For the medium educated workers, there is also a decrement in the years in formality, half a year lower on average. This happens only in the low part of the distribution but the medium and high parts of the distribution stay in the same shape as the benchmark.

Finally, the fourth policy is designed to put stricter requirements on the minimum age to go in retirement, 67 and 72 instead to 65 and 70 for the full and *Advance age*

	Low Education	
	Benchmark	Policy
Years of formality (at 65)	24.8	25.5
Years of formality (at 70)	27.3	28.5
	Distribution at age 70	
$p_{0.025}$	18	17
$p_{0.125}$	22	22
$p_{0.25}$	25	25
$p_{0.5}$	29	29
$p_{0.75}$	30	33
$p_{0.875}$	32	35
$p_{0.975}$	35	35
	Medium education	
	Benchmark	Policy
Years of formality (at 65)	34.0	34.9
Years of formality (at 70)	34.1	35.7
	Distribution at age 70	
$p_{0.025}$	29	33
$p_{0.125}$	31	35
$p_{0.25}$	32	35
$p_{0.5}$	34	35
$p_{0.75}$	36	36
$p_{0.875}$	37	37
$p_{0.975}$	38	39

Table 3.20: Formality with the second policy experiment: formality years of requirement 20 and 35.

	Full_3		Full_2		Full_1	
	Benchmark	Policy	Benchmark	Policy	Benchmark	Policy
Low Education	0.0%	0.0%	1.4%	0.4%	44.4%	76.2%
Medium Education	0.3%	0.2%	40.9%	28.9%	58.5%	71.0%
High Education	6.2%	3.3%	56.4%	51.5%	37.3%	45.5%
	Advanced Age		Survival			
	Benchmark	Policy	Benchmark	Policy		
Low Education	54.2%	23.44%	0.0%	0.0%		
Medium Education	0.3%	0.0%	0.0%	0.0%		
High Education	0.0%	0.0%	0.0%	0.0%		

Table 3.21: Pension achievement with the third policy experiment: formality years of requirement 25. Full_3 pension is 35 and 40 years, Full_2 pension is 30 and 35 years and Full_1 pension is 25 and 30 years in the policy and in the benchmark respectively.

	Low Education	
	Benchmark	Policy
Years of formality (at 65)	24.8	24.6
Years of formality (at 70)	27.3	25.9
	Distribution at age 70	
$p_{0.025}$	18	17
$p_{0.125}$	22	22
$p_{0.25}$	25	25
$p_{0.5}$	29	25
$p_{0.75}$	30	28
$p_{0.875}$	31	31
$p_{0.975}$	34	33
	Medium education	
	Benchmark	Policy
Years of formality (at 65)	34.0	33.3
Years of formality (at 70)	34.1	33.3
	Distribution at age 70	
$p_{0.025}$	29	27
$p_{0.125}$	31	30
$p_{0.25}$	32	32
$p_{0.5}$	34	33
$p_{0.75}$	36	35
$p_{0.875}$	37	36
$p_{0.975}$	39	38

Table 3.22: Formality with the third policy experiment: formality years of requirement 25.

pension respectively. This policy experiment is considered to obtain more resources to cover deficits[21]. Workers in all educative levels work more formally than in the benchmark after the age of 45 as is shown in Table E.6. Note for example, that the low educated work at the age of 67, 2.5 years more in formality than in the benchmark at the age of 65 (Table 3.24).

Pension achievement is shown in Table 3.23, workers in all educative levels achieve higher pensions due to the fact that they work more formally. However, 42% of the low educated only achieve an *Advance age* pension. The distribution of total years in formality (Table 3.24) is more variable for the low educated than in the benchmark.

[21]Note if most of the low educated workers continue working after the age of 65, then this requirement will not affect them. The budget could be affected if more workers have access to a full pension with higher replacement rates, but those parameters do not deeply affect the worker's behavior.

	Full_3		Full_2		Full_1	
	Benchmark	Policy	Benchmark	Policy	Benchmark	Policy
Low Education	0.0%	0.0%	1.4%	7.1%	44.4%	50.8%
Medium Education	0.3%	7.3%	40.9%	60.7%	58.5%	32.2%
High Education	6.2%	29.2%	56.4%	57.7%	37.3%	13.2%
	Advanced Age		Survival			
	Benchmark	Policy	Benchmark	Policy		
Low Education	54.2%	42.1%	0.0%	0.0%		
Medium Education	0.3%	0.0%	0.0%	0.0%		
High Education	0.0%	0.0%	0.0%	0.0%		

Table 3.23: Pension achievement with the fourth policy experiment: age of 67 as full pension requirement.

	Low Education	
	Benchmark	Policy
Years of formality (at 65/67)	24.8	27.2
Years of formality (at 70/72)	27.3	28.9
	Distribution at age 70	
$p_{0.025}$	18	19
$p_{0.125}$	22	24
$p_{0.25}$	25	27
$p_{0.5}$	29	30
$p_{0.75}$	30	31
$p_{0.875}$	31	33
$p_{0.975}$	34	37
	Medium education	
	Benchmark	Policy
Years of formality (at 65/67)	34.0	35.0
Years of formality (at 70/72)	34.1	35.0
	Distribution at age 70	
$p_{0.025}$	29	31
$p_{0.125}$	31	32
$p_{0.25}$	32	34
$p_{0.5}$	34	36
$p_{0.75}$	36	38
$p_{0.875}$	37	39
$p_{0.975}$	39	41

Table 3.24: Formality with the fourth policy experiment: age of 67 as full pension requirement.

3.8 Concluding Remarks

In this paper, I simulate a discrete choice model where the workers choose between working in the formal or in the informal sector. These choices depend on the wages that are paid in each sector (by education and experience), and the benefits that the workers can enjoy in retirement age. The model replicates the existence of a large share of informal salaried workers, especially in the lower side of the distribution where the low educated workers account for the majority.

The Argentinean pension system is considered strict in term of requirements in order to be covered at retirement age, which is a more significant problem for the low educated workers. However, the requirements to have the right to an *Advanced age* pension are not so strict for a salaried worker (only 10 years) and, both in my estimations and in the literature, all workers at least achieve this pension[22].

The model estimation captures the informality in all the educative levels and also the transitions among the formality, informality and unemployment. Pension achievements are in line with other estimations in the literature, showing that among the salaried workers the main problem is with the low educated ones.

The policy experiments show that workers decide to work less in formality when the compulsory career is shorter and replacement rates are lower (PAYG scheme), even if they achieve a lower pension income. When the requirements became stricter, they work more in formality, especially those workers who are closer to the new thresholds. It is remarkable that when the first step of full pension requirement is less strict, they work less formally, even those who are at the top of the distribution. The main changes in the behavior occur in the last part of the career, that is after the age of 45.

[22]Note, that this special pension was designed mainly for women and self employed workers.

3.9 Appendix

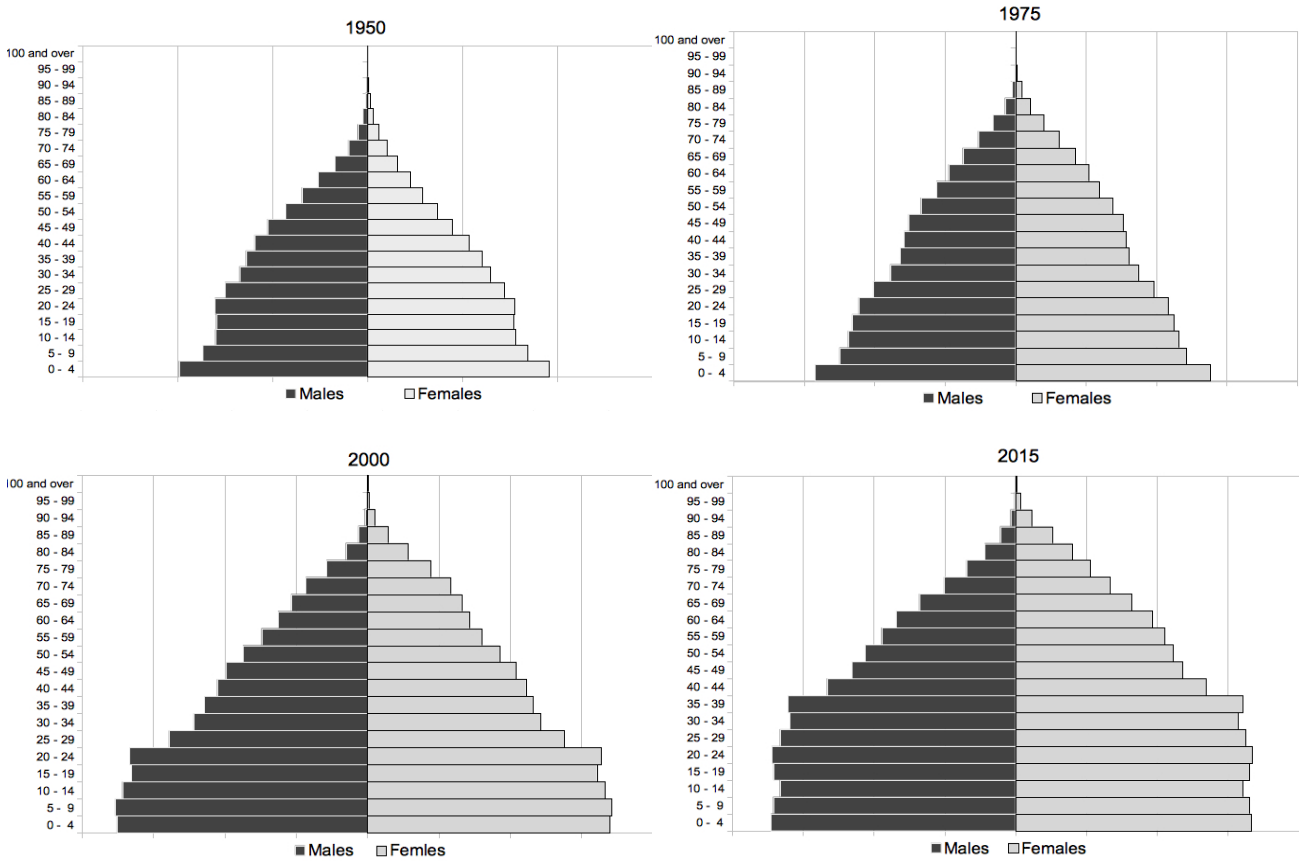
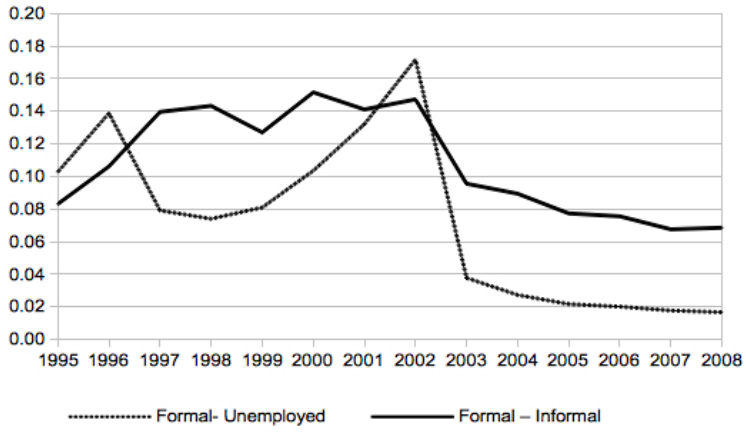
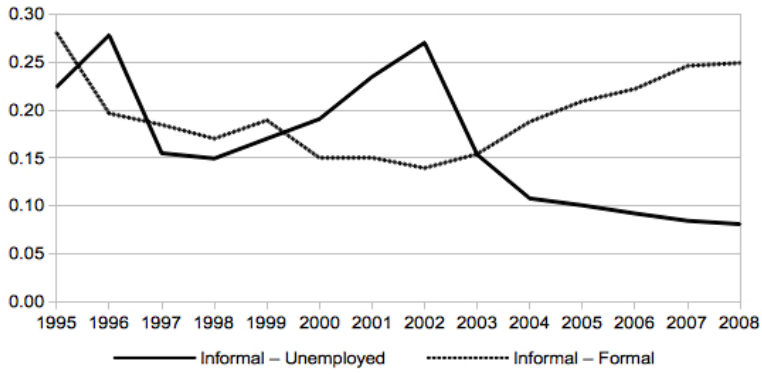


Figure E.1: Argentinean population pyramid 1995, 1975, 2000 and 2015.
Source: CELADE.

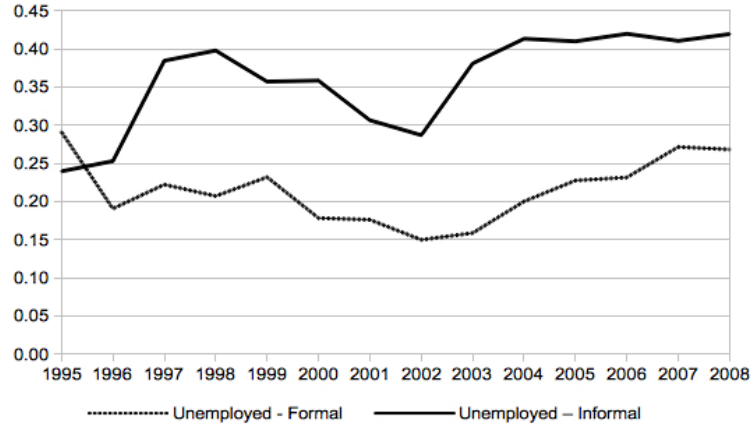


(a) From formal

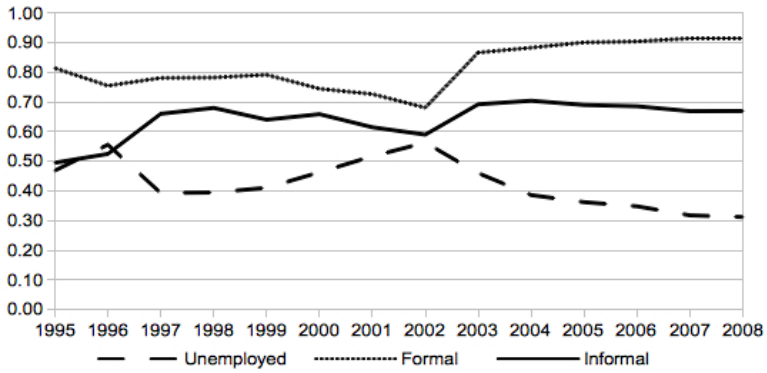


(b) From informal

Figure E.2: Transition from formality (informality) to the other states.



(a) From unemployment



(b) Stayer

Figure E.3: Transition from unemployment to the other states and those who remain in the same state.

Age group: 23-28 years old			
Unemployed	Unemployed (-1)	Formal (-1)	Informal (-1)
Low Education	0.3805	0.0597	0.1079
Medium Education	0.3741	0.0371	0.1094
High Education	0.3889	0.0329	0.1416
Formal	Unemployed (-1)	Formal (-1)	Informal (-1)
Low Education	0.1531	0.7707	0.1402
Medium Education	0.2291	0.8636	0.2276
High Education	0.3220	0.9086	0.3241
Informal	Unemployed (-1)	Formal (-1)	Informal (-1)
Low Education	0.4664	0.1695	0.7519
Medium Education	0.3969	0.0993	0.6630
High Education	0.2891	0.0584	0.534
Age group: 29-34 years old			
Unemployed	Unemployed (-1)	Formal (-1)	Informal (-1)
Low Education	0.3726	0.0413	0.0941
Medium Education	0.3533	0.0247	0.1021
High Education	0.3255	0.0151	0.1010
Formal	Unemployed (-1)	Formal (-1)	Informal (-1)
Low Education	0.1781	0.8215	0.1733
Medium Education	0.2723	0.8992	0.2671
High Education	0.3902	0.9435	0.4046
Informal	Unemployed (-1)	Formal (-1)	Informal (-1)
Low Education	0.4493	0.1372	0.7326
Medium Education	0.3744	0.0762	0.6308
High Education	0.2843	0.0415	0.4944
Age group: 35-44 years old			
Unemployed	Unemployed (-1)	Formal (-1)	Informal (-1)
Low Education	0.3736	0.0356	0.1039
Medium Education	0.3480	0.0199	0.1043
High Education	0.2972	0.0104	0.0977
Formal	Unemployed (-1)	Formal (-1)	Informal (-1)
Low Education	0.2003	0.8471	0.1888
Medium Education	0.3055	0.9171	0.3058
High Education	0.4405	0.9560	0.4438
Informal	Unemployed (-1)	Formal (-1)	Informal (-1)
Low Education	0.4262	0.1173	0.7073
Medium Education	0.3466	0.0630	0.5899
High Education	0.2623	0.0336	0.4585

Table E.1: Probabilities of being in each sector (only men) based on the multinomial model.

Age group: 45-54 years old			
Unemployed	Unemployed (-1)	Formal (-1)	Informal (-1)
Low Education	0.3829	0.0330	0.1143
Medium Education	0.3352	0.0179	0.1152
High Education	0.2999	0.0094	0.0936
Formal	Unemployed (-1)	Formal (-1)	Informal (-1)
Low Education	0.2220	0.8654	0.2072
Medium Education	0.3450	0.9275	0.3289
High Education	0.4725	0.9616	0.4876
Informal	Unemployed (-1)	Formal (-1)	Informal (-1)
Low Education	0.3951	0.1016	0.6786
Medium Education	0.3198	0.0546	0.5559
High Education	0.2277	0.0291	0.4188
Age group: 55-65 years old			
Unemployed	Unemployed (-1)	Formal (-1)	Informal (-1)
Low Education	0.4235	0.0379	0.1351
Medium Education	0.3568	0.0209	0.1395
High Education	0.2624	0.0092	0.0864
Formal	Unemployed (-1)	Formal (-1)	Informal (-1)
Low Education	0.2245	0.8688	0.2226
Medium Education	0.3605	0.9290	0.3437
High Education	0.5372	0.9654	0.5351
Informal	Unemployed (-1)	Formal (-1)	Informal (-1)
Low Education	0.3520	0.0933	0.6423
Medium Education	0.2828	0.0501	0.5168
High Education	0.2003	0.0253	0.3785

Table E.2: Probabilities of being in each sector (only men) based on the multinomial model.

	Low Education		Medium Education		High Education	
	Data	Model	Data	Model	Data	Model
23-28	0.1462	0.0862	0.0830	0.1039	0.0808	0.0661
	[0.1298, 0.1703]		[0.0694, 0.0966]		[0.0464, 0.1147]	
29-34	0.0845	0.0362	0.0527	0.0326	0.0172	0.0056
	[0.0710, 0.0968]		[0.0384, 0.0675]		[0.0090, 0.0429]	
35-44	0.0729	0.0342	0.0311	0.0032	0.0126	0.0034
	[0.0618, 0.0861]		[0.0232, 0.0458]		[0.0076, 0.0222]	
45-54	0.1053	0.1091	0.0220	0.0031	0.0161	0.0355
	[0.0696, 0.1551]		[0.0152, 0.0354]		[0.0108, 0.0227]	
55+	0.0862	0.0970	0.1044	0.0925	0.0096	0.0307
	[0.0624, 0.1122]		[0.0514, 0.1687]		[0.0047, 0.0174]	

Table E.3: Moment matching: Unemployed workers. The interval is estimated based on the multinomial model, performing bootstraps and considering the middle 90%.

	General		Low Education		Medium Education		High Education	
	Benchmark	Policy	Benchmark	Policy	Benchmark	Policy	Benchmark	Policy
23-28	0.399	0.398	0.265	0.264	0.481	0.480	0.615	0.612
29-34	0.619	0.622	0.512	0.513	0.692	0.697	0.778	0.780
35-44	0.756	0.746	0.651	0.628	0.852	0.851	0.854	0.857
45-54	0.747	0.757	0.544	0.550	0.943	0.955	0.904	0.924
55+	0.831	0.891	0.764	0.838	0.875	0.937	0.933	0.949

Table E.4: Formality path with the second policy experiment: 35 years of formality as full pension requirement and 20 years in formality as *Advance age* pension requirement.

	General		Low Education		Medium Education		High Education	
	Benchmark	Policy	Benchmark	Policy	Benchmark	Policy	Benchmark	Policy
23-28	0.399	0.398	0.265	0.264	0.481	0.484	0.615	0.619
29-34	0.619	0.620	0.512	0.513	0.692	0.694	0.778	0.780
35-44	0.756	0.758	0.651	0.628	0.852	0.853	0.854	0.854
45-54	0.747	0.720	0.544	0.544	0.943	0.894	0.904	0.848
55+	0.831	0.807	0.764	0.743	0.875	0.839	0.933	0.922

Table E.5: Formality path with the third policy experiment: 25 years of formality as full pension requirement.

	General		Low Education		Medium Education		High Education	
	Benchmark	Policy	Benchmark	Policy	Benchmark	Policy	Benchmark	Policy
23-28	0.399	0.402	0.265	0.279	0.481	0.484	0.615	0.606
29-34	0.619	0.624	0.512	0.513	0.692	0.699	0.778	0.785
35-44	0.756	0.753	0.651	0.647	0.852	0.849	0.854	0.854
45-54	0.747	0.749	0.544	0.533	0.943	0.948	0.904	0.939
55+	0.831	0.889	0.764	0.854	0.875	0.906	0.933	0.955

Table E.6: Formality path with the fourth policy experiment: age of 67 as full pension requirement.

	Transitions				
	Benchmark	1st Policy	2nd Policy	3rd Policy	4th Policy
Informal-Formal	0.476	0.485	0.484	0.465	0.474
Informal-Unemployed	0.123	0.113	0.100	0.138	0.122
Informal-Informal	0.401	0.402	0.416	0.397	0.404
Formal-Formal	0.835	0.795	0.85	0.825	0.853
Formal-Unemployed	0.027	0.036	0.014	0.033	0.019
Formal-Informal	0.139	0.166	0.136	0.142	0.129
Unemployed-Formal	0.259	0.347	0.224	0.349	0.234
Unemployed-Unemployed	0.323	0.218	0.223	0.292	0.292
Unemployed-Informal	0.418	0.435	0.553	0.359	0.475

Table E.7: Transitions with the policy experiments.

Bibliography

Alexander, K., Entwisle, D., and Kabbani, N. (2001). The dropout process in life course perspective: Early risk factors at home and school. *Teachers College Record*, 103:760–823.

Alm, J., Jackson, B. R., and McKee, M. (2009). Getting the word out: Enforcement information dissemination and compliance behavior. *Journal of Public Economics*, 93(3-4):392–402.

Amaral, P. S. and Quintin, E. (2006). A competitive model of the informal sector. *Journal of Monetary Economics*, 53(7):1541–1553.

Arza, C. (2009). Back to the state: Pension fund nationalization in argentina. Technical Report DT 73, Centro Interdisciplinario para el Estudio de Políticas Públicas.

Attanasio, O., Meghir, C., and Santiago, A. (2010). Education choices in mexico: using a structural model and a randomized experiment to evaluate progresa. IFS Working Papers W10/14, Institute for Fiscal Studies.

Bailey Jones, J. and French, E. (2010). The effects of health insurance and self-insurance on retirement behavior. Technical report, University at Albany, SUNY, Department of Economics.

Bertranou, F. (2001). *Cobertura Previsional en Argentina, Brasil y Chile*. Oficina Internacional del Trabajo.

Boca, D. D., Monfardini, C., and Nicoletti, C. (2012). Children's and parents' time-use choices and cognitive development during adolescence. Working Papers 2012-006, Human Capital and Economic Opportunity Working Group.

Bosch, M. and Guajardo, J. (2012). Labor market impacts of non-contributory pensions: The case of argentina's moratorium. IDB Publications 78158, Inter-American Development Bank.

Bosch, M. and Manacorda, M. (2012). Social policies and labor market outcomes in latin america and the caribbean: A review of the existing evidence. CEP Occasional Papers 32, Centre for Economic Performance, LSE.

Botero García, J. (2010). El impacto de las restricciones a las exportaciones colombianas a venezuela. CIRJE F-Series 10-04, Centro de Investigaciones Económicas y Financieras.

Browning, M., Chiappori, P., and Weiss, Y. (2014). *Economics of the Family*. Cambridge Surveys of Economic Literature. Cambridge University Press.

Bucheli, M. and Ceni, R. (2010). Informality sectoral selection and earnings in uruguay. *Estudios Económicos*, 25(2):281–307.

Cappellari, L., Dell'Aringa, C., and Leonardi, M. (2011). Temporary employment, job flows and productivity: A tale of two reforms. Working paper series.

Christle, C., Jolivette, K., and Nelson, C. (2002). The high school journal. *Remedial and Special Education*, 85(4):325–339.

Dolado, J. J., Garca-Serrano, C., and Jimeno, J. F. (2001). Drawing lessons from the boom of temporary jobs in spain. Working Papers 2001-11, FEDEA.

Eckstein, Z. and Wolpin, K. I. (1999). Why youths drop out of high school: The impact of preferences, opportunities, and abilities. *Econometrica*, 67(6):1295–1340.

Fields, G. S. (2011). Labor market policy in developing countries : a selective review of the literature and needs for the future. *Labour Economics*, 18.

Finn, J. (1989). Withdrawing from school. *Review of Educational Research*, 59(2):117–142.

Forteza, A., Apella, I., Fajnzylber, E., Grushka, C., Rossi, I., and Sanroman, G. (2009). Work histories and pension entitlements in argentina, chile and uruguay. Social Protection Discussion Papers 52446, The World Bank.

Forteza, A. and Ourens, G. (2009). How much do latinamerican promise to pay back? Technical Report 31/09, dECON.

Friedman, E., Johnson, S., Kaufmann, D., and Zoido-Lobaton, P. (2000). Dodging the grabbing hand: the determinants of unofficial activity in 69 countries. *Journal of Public Economics*, 76(3):459–493.

Galiani, S. and Weinschelbaum, F. (2007). Modeling informality formally: Households and firms. CEDLAS, Working Papers 0047, CEDLAS, Universidad Nacional de La Plata.

Giuliodori, D. and Stucchi, R. (2010). Innovation and job creation in a dual labor market: evidence from spain. MPRA Paper 23006, University Library of Munich, Germany.

Griffin, B. (2002). Academic disidentification, race, and high school dropouts. *High School Journal*, 85(4):71–81.

Hazans, M. (2011). Informal workers across europe: Evidence from 30 countries. IZA Discussion Papers 5871, Institute for the Study of Labor (IZA).

Holzmann, R; Robalino, D. and Takayama, N. e. (2009). *Closing the Coverage Gap: The Role of Social Pensions and Other Retirement Income Transfers*. World Bank, robert holzmann, david a. robalino, noriyuki takayama edition.

Ihrig, J. and Moe, K. S. (2004). Lurking in the shadows: the informal sector and government policy. *Journal of Development Economics*, 73(2):541–557.

ILO (2013). Trabajo decente y juventud en américa latina. 2013. World bank other operational studies, Oficina Regional para América Latina y el Caribe.

Johnson, S., Kaufmann, D., and Zoido-Lobaton, P. (1998). Regulatory discretion and the unofficial economy. *American Economic Review*, 88(2):387–92.

Joubert, C. (2012). Pension design with a large informal labor market: Evidence from chile. Technical report, UPenn.

Kaufmann, K. M., Ferrara, E. L., and Brollo, F. (2012). Learning about the enforcement of conditional welfare programs: Evidence from the bolsa familia program in brazil. Technical report, Department of Economics, Bocconi University.

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

Kmenta, J. (1967). On estimation of the ces production function. *International Economic Review*, 8:180–189.

Labat, J. P. (2012). La perspectiva del mides en investigación y políticas sociales. La colaboración entre la udelar y el mides para la implementación del panes, CSIC-UdelaR.

Li, Y. and Mumford, K. (2009). Aspirations, expectations and education outcomes for children in britain: Considering relative measures of family efficiency. Discussion Papers 09/26, Department of Economics, University of York.

Loayza, N. V. and Rigolini, J. (2011). Informal employment: Safety net or growth engine? *World Development*, 39(9):1503–1515.

Loayza, N. V., Serven, L., and Sugawara, N. (2009). Informality in latin america and the caribbean. Policy Research Working Paper Series 4888, The World Bank.

Magnac, T. (1991). Segmented or competitive labor markets. *Econometrica*, 59(1):165–87.

Maloney, W. F. (2004). Informality revisited. *World Development*, 32(7):1159–1178.

Otero, A. (2013). Informality and pension incentives: A structural household life cycle model of consumption, labor supply and pension savings. Technical report, UCL.

Portes, A., Castells, M., Benton, L., and Portes, A. (1989). *The Informal economy : studies in advanced and less developed countries*. Baltimore, Md.: Johns Hopkins University Press, castells, m. and benton, l. (eds.) edition.

Pratap, S. and Quintin, E. (2006). Are labor markets segmented in developing countries? a semiparametric approach. *European Economic Review*, 50(7):1817–1841.

Rincke, J. and Traxler, C. (2011). Enforcement Spillovers. *The Review of Economics and Statistics*, 93(4):1224–1234.

Robalino, D.and Zylbertajn, H. (2009). Ex-ante methods to assess the impact of social insurance policies on labor supply with an application to brazil. Technical report, World Bank.

Rofman, R. (2000). The pension system in argentina six years after the reform. Technical report, World Bank.

Rofman, R., E., F., and Herrera, G. (2010). Reformando las reformas previsionales: en la argentina y chile. *Revista CEPAL*, 101:85–109.

Schneider, F. (2012). The shadow economy and work in the shadow: What do we (not) know? IZA Discussion Papers 6423, Institute for the Study of Labor (IZA).

Stinebrickner, T. and Stinebrickner, R. (2013). Academic performance and college dropout: Using longitudinal expectations data to estimate a learning model. Working Paper 18945, National Bureau of Economic Research.

Todd, P. and Velez-Grajales, V. (2008). How pension rules affect work and contribution patterns: a behavioral model of the chilean privitized pension system. Technical report, University of Michigan. Retirement Research Center.

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.

Ulyssea, G. (2010). Regulation of entry, labor market institutions and the informal sector. *Journal of Development Economics*, 91(1):87–99.

UM/CIEA (2013). Educación en uruguay y américa latina. Boletín estadístico, Centro de Investigaciones en Economía Aplicada, Universidad de Montevideo.

United-Nations (1999). World population aging 1950-2050. Technical report, United Nations, New York, NY. Dept. of Economic and Social Affairs.

Van der Klaauw, W. and Wolpin, K. I. (2008). Social security and the retirement and savings behavior of low-income households. *Journal of Econometrics*, 145:21–42.

World-Bank (2010). Informality in colombia : Implications for worker welfare and firm productivity. World Bank Other Operational Studies 2889, The World Bank.