



# Expectations Formation and Optimal Taxation

Luis E. Rojas

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

Florence, 18 November 2016



European University Institute  
**Department of Economics**

## Expectations Formation and Optimal Taxation

Luis E. Rojas

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

### **Examining Board**

Prof. Ramon Marimon, EUI, Supervisor  
Prof. Árpád Ábrahám, EUI  
Prof. Klaus Adams, University of Mannheim  
Prof. George Evans, University of Oregon

© Rojas, 2016

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



# Abstract

This thesis is composed of three essays that propose macroeconomic theories to answer empirical questions and guide policy design. The focus is on expectations formation, learning and optimal taxation.

In the first chapter I address the empirical finding that sovereign default history is a predictor of risk spreads even after controlling for pricing fundamentals. I show that this fact can be reconciled with a model of creditors' learning about the default probability of a sovereign. Furthermore, I show that if creditors learn about a group of countries, then clusters of default emerge as a side effect of the beliefs formation process.

The second chapter documents that investment recovers sluggishly after recessions and that consumption tends to lead the recoveries. I propose a model to show that during recessions investors might be excessively pessimistic about consumer demand and delay the implementation of projects. Taking this setting to the design of countercyclical policy, I argue that corporate income taxation can be a desirable instrument to use, as it is linked to the expected gains of firms, while interest rate policy or investment subsidies affect the cost of investment.

In the third chapter, coauthored with Paweł Doligalski, we study how to design the tax system in an economy featuring an informal labor market, by extending the Mirrlees theory on optimal income taxation. We estimate the key elements of this model on Colombian data and compute the optimal tax schedule.

# Acknowledgments

The completion of this thesis was a challenging and exciting journey that I finish highly indebted to many people. I will briefly mention just a few names but I am extremely fortunate as to have many more in my mind. If you are reading this you are most likely one of them, so THANK YOU!

I thank my supervisor Ramon Marimon for his continuous support and encouragement. I particularly value his well aimed criticism that pushed me to realize the large difference between an idea and a piece of research. This thesis is an ongoing effort to get there. I would also like to thank Árpád Ábrahám, Evi Pappa, Fabio Canova and Juan J. Dolado. I highly benefited of the many discussions with you, from which I draw many lessons.

My fellow PhD students were crucial for the completion of this thesis, the many comments and suggestions I got about my work and the many things I learned from theirs is something I deeply value. In particular I thank my coauthor Paweł Doligalski. Also, for showing me two completely opposite but equally exciting ways to do research I thank Nacho Gonzalez and Pavel Brendler.

Life in Florence was simply full of joy. A big credit for that goes to GaBo, Lian & Nico “i boludi”, los vag@s, Sonia+Lori, my cohort, IUE calcio and last but most importantly Zelda!

I thank the EUI and the staff at the Econ department. You have created a great institution and I feel extremely lucky that I got the chance to do my PhD here.

Last, I thank Gabo, Mao, Mimi and Luis Eduardo for all the summer road trips :P

*To someone that has always been  
there,*

*Abel*

*and to someone that is about to  
come,*

*Emma/Gabriel*



# Contents

## Chapter 1

Learning In Sovereign Debt Markets.....	1
1. Introduction .....	2
2. Reaction of spreads to default and related literature.....	4
3. Model .....	6
3.1. The government	6
3.2. The creditors and the equilibrium bond price menu	9
4. Spreads after default.....	12
5. Clusters of default.....	15
6. Discussion of results, caveats and alternative theories.....	17
7. Conclusions.....	19
References .....	21
A. Appendix .....	23
A.1. Perceived law of motion L	25

## Chapter 2

Expectations Formation and Investment During Recessions.....	27
1. Introduction .....	28
2. Investment recovery after a recession.....	29
3. Related literature .....	31
4. The basic model.....	34
4.1. Setup	36
4.2. Equilibrium definitions	39
4.3. Subjective beliefs	40
4.4. Equilibrium	43
4.5. Objective beliefs	44
4.6. Subjective beliefs	44
4.7. Extension to a T-Period economy	47
5. Full model.....	51
5.1. Setup	51

5.2. Equilibrium dynamics and impulse responses	56
6. Policy interventions.....	60
6.1. Corporate income tax	60
6.2. Comparison of policies	63
6.3. Recession caused by a productivity shock	65
6.4. Delay in investment driven by technological risk	66
7. Conclusions and caveats.....	70
References .....	71

### Chapter 3

Optimal Redistribution with a Shadow Economy .....	73
1. Introduction .....	74
2. Simple model .....	77
2.1. First-best	78
2.2. Second-best	79
3. Full model .....	85
3.1. The planner's problem	86
3.2. Incentive compatibility	87
3.3. Optimality conditions	88
3.4. Summary of results	97
4. Measuring shadow and formal productivities .....	99
4.1. Data	100
4.2. Empirical specification	104
4.3. Estimation results	106
5. Calibrated exercise.....	107
5.1. Calibration of the utility function	107
5.2. Optimal allocations	108
6. Conclusions .....	111
References .....	112
A. Appendix: Proofs .....	113
A.1. Proofs from section 2	113
A.2. Proofs from section 3	119
B. Appendix: Estimation.....	126

# Chapter 1

## Learning in Sovereign Debt Markets

### **Abstract**

A default episode is typically followed by 1) a raise in spreads and 2) a positive surplus for the lenders. In this paper I show that this fact can be reconciled with a model of creditors learning about the default probability of a sovereign. Existing sovereign debt models are instead unable to accommodate these aspects of the data, as they portray mappings from economy fundamentals to spreads which are not affected by default history. The theory also implies that full repayment of debt can lead to a drop in spreads, providing the incentives to the sovereign for higher debt exposure and consequently a negative surplus for the lenders. Furthermore, if creditors learn about multiple countries, I show that clusters of default emerge as a side effect of the beliefs formation process, and can occur even if there is no trade or capital market link between countries.

# 1 Introduction

Sovereign debt has historically been one of the largest classes of financial assets. In 1950 it accounted for 22% of the market value of worldwide assets, and for 19% in 2010.<sup>1</sup> It is also one of the main instruments for governments to finance their expenditures. Therefore, it is relevant to understand how the market prices these assets.

According to the Standard & Poor's default criterion<sup>2</sup>, more than 240 sovereign defaults occurred over the period 1824-2004. This makes of default a key element affecting the sovereign debt market, and consequently the pricing of this asset. In fact, the probability of default has been proposed as one of the main factors accounting for the interest rate premia that sovereign debt pays over a "risk free" asset (commonly known as spreads).<sup>3</sup>

Empirical studies have found that the default history is informative about the variation of spreads. This finding holds even after controlling for the variables that economic theory identifies as the key determinants of default decisions, the so-called "fundamentals". Two facts highlighted in this literature are: a default episode is typically followed by 1) a raise in spreads and 2) a positive surplus for the lenders. The main contribution of this paper is to provide a theory that can account for both facts.

The first fact is documented in Cruces and Trebesch (2013), who, using a panel of countries for the period 1970-2010, estimate that an increase of the haircut<sup>4</sup> by one percentage point generates an increase of 4-5 basis points in spreads 4 to 7 years after the settlement. This pattern has also been found in other analyses on risk spreads and credit ratings (see Borensztein and Panizza (2009), Cantor and Packer (1996), Afonso, Gomes, and Rother (2006) and Panizza, Sturzenegger, and Zettelmeyer (2009)). The second fact is documented in Benczur and Ilut (2016), where, using a panel of countries for the period 1970-1982, the authors show that the raise in spreads after default is higher than what is implied by country fundamentals, investors risk aversion and the objective default probability.

These two facts are not accounted for by the existing models of sovereign debt. Sovereign debt models are characterized by a rational expectations equilibrium (REE) where, for each possible value of fundamentals and debt issuance, there is a corresponding spread that reflects the default probabilities, such that creditors break even on average<sup>5</sup>. In most of these models the default event is not informative and therefore spreads do not react to a default episode.<sup>6</sup> The exceptions

<sup>1</sup>During the 19th Century, sovereign debt even reached a percentage higher than 70% of all assets traded in the London Stock Exchange. See Tomz and Wright (2013).

<sup>2</sup>"...the failure to meet a principal or interest payment on the due date (or within the specified grace period) contained in the original terms of a debt issue ... or tenders an exchange offer of new debt with less-favorable terms than the original issue" (Beers and Cavanaugh, 2006) taken from Richmond and Dias (2008). See Borensztein and Panizza (2009) for a discussion on others criteria to classify default.

<sup>3</sup>See the surveys presented in Tomz and Wright (2013) and Aguiar and Amador (2014). The quantitative sovereign debt models in Aguiar and Gopinath (2006); Arellano (2008) are an example of this. The early theoretical contribution of Grossman and Van Huyck (1988) lays out the theoretical the framework to link default and spreads in sovereign debt markets.

<sup>4</sup>The percentage loss of the asset value for the creditor due to a renegotiation of the outstanding debt.

<sup>5</sup>This mapping is commonly named the 'bond price menu'.

<sup>6</sup>See for example Arellano (2008); Yue (2010).

are Bayesian REE, where a default decision can be informative and the spread may change after a default episode, due to asymmetric information between creditors and government. Nevertheless, as creditors still break even on average, these models cannot account for the emergence of a positive surplus for creditors. I also question the assumption that creditors possess detailed and exact knowledge about governments, which is imposed in these modeling choices.

I propose a sovereign debt model that can account for both facts through a minor deviation from rational expectations. Creditors are endowed with a well-specified set of beliefs about the sovereign decision to default. These beliefs might not coincide exactly with REE beliefs. Given those beliefs, creditors participate in sovereign debt markets and behave rationally.<sup>7</sup> This is the concept of Internal Rationality proposed by Adam and Marcet (2011).

While the REE imposes a unique mapping from fundamentals to market outcomes, the belief system of the creditors in this model allows for a joint probability distribution of market outcomes (default or repayment) and fundamentals. This design accounts for the fact that agents might not have perfect market knowledge, in particular that creditors have uncertainty about the sovereign and about the political decision process that leads to repayment or default.<sup>8</sup> The deviation from the REE is minor, as the creditors' system of beliefs is centered around the REE beliefs and can be set arbitrarily close to it.

The key mechanism of this paper is: whenever creditors face a higher (lower) haircut than expected, they update their beliefs about default probabilities and, keeping fundamentals unchanged, spreads rise (fall). Therefore, the spreads increase after a default and might do so beyond what is implied by the 'objective' default probability, allowing for a positive surplus for creditors. The updating algorithm implied by the belief system of the creditors is a variant of a fixed-gain learning algorithm.<sup>9</sup>

Just as the model replicates the rise in spreads and a possible surplus after default, the learning mechanism embeds a certain degree of symmetry and therefore repayment is also informative and leads to lower spreads and a possible negative surplus for creditors. Altogether spreads fluctuate around the REE case, in which the endogenous mean reversion mechanism is the decision of the sovereign to default or repay and debt issuance. I further show that the learning mechanism can account for clusters of default, when creditors attribute common features to a group of countries.

The rest of the paper is organized as follows: first, I discuss the highlighted stylized facts and the related literature, then I present the model, the extension and a discussion of the results and of alternative theories. I conclude pointing out the insights and caveats of the paper.

---

<sup>7</sup>Taking their belief system as given they maximize the expected discount stream of payoffs and have dynamically consistent plans.

<sup>8</sup>As opposed to the Bayesian RE creditors that knows perfectly all details of the decision making process of the government but just cannot observe perfectly the values of some parameters that are part of the problem.

<sup>9</sup>See Evans and Honkapohja (2001) for a detailed exposition of learning algorithms.

## 2 Reaction of spreads to default and related literature

It is well documented that the default history of a sovereign can partly account for the variation in sovereign debt spreads across countries and time.<sup>10</sup> The consensus in the literature is that sovereigns with a record of defaults pay higher spreads than those who have not incurred in one (see Tomz (2007), Flandreau and Zumer (2003), Cantor and Packer (1996), Afonso, Gomes, and Rother (2006) and Panizza, Sturzenegger, and Zettelmeyer (2009)). Nevertheless, the overall effects seem to be small and short lived. Borensztein and Panizza (2009) estimate a panel regression model for 31 emerging markets for the sample 1997-2004 and find that spreads rise after a default but the effects are short-lived. They control for a set of variables that are considered “fundamentals” of the sovereign bond pricing, such as GDP growth, the level of debt and the current account.

On the other hand, Cruces and Trebesch (2013) find that the effects can be sizable and economically significant at least seven years after the settlement is reached. The main difference in Cruces and Trebesch (2013) with respect to previous literature is a more precise measurement of creditors losses during a default. In particular, they explicitly take into account the maturity structure of the defaulted debt and the newly issued debt included in the settlement<sup>11</sup>. Their sample spans over a panel of 68 countries for the period 1970-2010; where 180 debt restructurings occurred. Their main result is that defaults having a haircut one standard deviation higher than the average, are associated spreads that are 120 basis points higher 4 to 7 years after the settlement.

Using a sample for 1970-1982 Benczúr and Ilut (2016) find that the raise in spreads after default is higher than what can be explained by a higher risk aversion or the objective default probability. They use the Error-in-Variables method (EVM) to recover a measure of the objective default probability.<sup>12</sup>

The model I propose provides a formal framework to the concepts discussed in Tomz (2007) of “surprising payers” and “expected defaulters”. In line with the evidence presented by Tomz (2007), the model prescribes that if the country chooses a haircut smaller than the perceived default probability, then creditors update their beliefs and spreads fall. If the country defaults and sets a haircut equal to the perceived default probability, then spreads do not react.

A closely related notion are also the “unjustifiable” and the “excusable” haircuts presented in Grossman and Van Huyck (1988). The authors construct a reputational equilibrium where for each contingency there is an expected haircut by the government, and this is properly accounted for in the pricing of sovereign debt. The model I present here generalizes the idea of Grossman and Van Huyck (1988). In their model, the after-default spreads are infinite in the case of an “unjustifiable” haircut; and the spreads remain unchanged if the haircut is exactly equal to the excusable (expected) level. In the model presented here there is a smoother relationship between the level

<sup>10</sup>See Tomz and Wright (2013) for a survey on the empirical research on sovereign debt markets.

<sup>11</sup>It is common that debt restructuring deals include new debt contracts

<sup>12</sup>The identification strategy consists in using realized repayment in the regression equation to have the expectational error as part of the error in the regression. Then, by a rational expectations argument they instrument the regression with variables known before the repayment was realized.

of the haircut and the spreads; the threshold of Grossman and Van Huyck (1988) represents the particular case in which creditors discount past information quickly enough.

The incomplete markets models of sovereign debt as Aguiar and Gopinath (2006) and Arellano (2008)<sup>13</sup> sustain debt in equilibrium by temporary exclusion from debt markets and lower realizations of the GDP during the default episode. In these models a key element of analysis is the bond-price menu. The bond price menu is a mapping from the economy fundamentals, such as the level of GDP and outstanding debt, to the spreads payable at each level of debt. This REE requires this menu to reflect the exact default probabilities of each sovereign debt contract, and the menu does not change after a default. Default history is not informative in this setup.

Alternatively, there are models that have considered information asymmetry between the country and the creditors: see for example Alfaro and Kanczuk (2005) and Catão, Fostel, and Kapur (2009). These models can generate a rise in spreads after default, as default can be informative about the government or the income process of the country (as in Catão, Fostel, and Kapur (2009)). Nevertheless, this approach cannot lead to a higher surplus for creditors after default, since the rise in spreads objectively reflect the higher default probability conditional on observables. Furthermore I will illustrate that the equilibrium concept adopted in this case demands the creditor to hold a large amount of information about the problem of the government and about the beliefs of all other creditors in the market.

Benczúr and Ilut (2016) interpret the higher surplus as evidence of a punishment by creditors on defaulters, and to be suggestive of a relational contract between the bank and the sovereign. This implies assumptions of either coordination or lock-in relationships between sovereigns and creditors. Nevertheless, coordination and lock-in relationships are not common in sovereign debt markets.<sup>14</sup>

The analysis in this paper uses the Internal Rationality concept of Adam and Marcet (2011) to study the pricing of sovereign debt. Adam, Marcet, and Nicolini (2016) have already exploited this concept to explain stylized facts in the asset pricing literature by slightly modifying the basic consumption based asset pricing model and relaxing the REE concept. They are able to generate, among other moments, the volatility and persistence of the price-dividend ratio for stocks.

In the setup by Adam, Marcet, and Nicolini (2016), the investors can trade an asset that has an exogenous stochastic process for dividends. In this case, a the rational expectations equilibrium would impose a mapping from the history of dividends to the equilibrium price, and consequently prices are not informative for the traders if they can observe dividends. They depart from rational expectations by allowing for the possibility that market participants do not know this mapping and form expectations with a system of beliefs that is characterized by a joint density of the history of dividends and prices The key point of Adam, Marcet, and Nicolini (2016) (henceforth AMN) is that if agents do not have perfect market knowledge, then expectations can be driven not only by beliefs about fundamentals (dividends in their case) but also market outcomes (prices). I follow

<sup>13</sup>Also known as the quantitative models of sovereign debt (See Aguiar and Amador (2014)).

<sup>14</sup>See Wright (2005) for a discussion about the not credible menace of creditors to punish a defaulting country. Also Kletzer and Wright (2000) presents the difficulties of imposing such penalties on sovereigns.

this notion and creditors in sovereign debt markets react not only to the fundamentals (GDP of the sovereign, level of debt, etc.) but also to market outcomes (default or repayment). The novelty of the analysis I present is that the decisions of the issuer of the asset are endogenous and depend on the belief system of investors.<sup>15</sup>

### 3 Model

The model economy is populated by a government and international creditors. The government is a benevolent planner of a small open economy, that collects taxes and issue bonds to finance expenditures. International creditors are risk neutral investors that value bonds according to their perceived default probability. In case of a default, the country is excluded from capital markets until a settlement with the creditors is achieved.

#### 3.1 The Government

The government maximizes the welfare of a small open economy through the provision of public goods. The preferences of the representative household are given by:

$$\sum_{t=0}^{\infty} \beta^t E \{u(c_t, g_t)\} \quad (1)$$

where  $c$  is private consumption;  $g$  is the public provision of goods;  $\beta$  the discount factor and  $u$  is a strictly concave and twice differentiable utility function.

At each period  $t$  the households receive an endowment  $y_t$  that follows a Markov chain with possible values  $Y = \{y^1, \dots, y^N\}$  that satisfy  $y^1 < y^2 < \dots < y^N$ . The transition probabilities are given by  $\pi(y' | y)$  for  $y', y \in Y$ . The households do not save and consume the after tax income such that  $c_t = (1 - \tau)y_t$ ; where  $\tau$  is the tax rate.

Every period the government has to cover an operational cost  $e_t$  that is i.i.d with mass probability function  $f(e)$  and domain  $E = \{e^1, e^2, \dots, e^M\}$ .<sup>16</sup> Let  $b_t$  be the amount of bonds issued in period  $t$  that mature in period  $t + 1$  and have face value equal to 1. Then, if the government fully repays the outstanding debt the budget constrain is

$$g_t + e_t - \tau y_t \leq q_t(b_t, y_t)b_t - b_{t-1}$$

---

<sup>15</sup>On the other hand the analysis is greatly simplified by the assumption of risk neutral investors.

<sup>16</sup>Another possible interpretation is to consider that, at least partly,  $e_t$  captures some rent extraction of the government from total resources. Along these lines, the government is not an ideal benevolent planner and we only require that at least the resources not appropriated by politicians (and operational costs) are used to maximize social welfare. I will not carry this interpretation along the paper because of the simple structure provided to the stochastic process  $\{e_t\}$ .

Quantitatively the inclusion of  $\{e_t\}$  is done to replicate that fact found in Tomz and Wright (2007) that although GDP is a strong predictor of default, the relation is weaker than what most models prescribe.



where  $q_t(b_t, y_t)$  is the market price of the bonds issued at time  $t$ . In general  $q_t(b, y)$  is a time-varying function of the amount of debt issued and income; commonly known in the literature as the bond-price menu. The constraint establishes that the primary deficit has to be lower or equal than net capital inflows.<sup>17</sup>

### Default decision, exclusion and the settlement

If the sovereign decides to default at period  $t$ , it is excluded from capital markets and the households suffer a  $(\alpha)$  percentage loss in their income every period until a settlement with the creditors is reached. The bargaining procedure for the settlement is as follows: every period, while the country is in default, the government can exit default if it offers to the creditors a haircut equal to  $\bar{h}$ ; the creditors reservation haircut level. The repayment can be done through immediate payments or handing newly issued bonds that creditors value at the market price.<sup>18</sup>

Let  $V(b_{t-1}, y_t, e_t, q_t(b, y))$  denote the value function of a government that has the option to default at  $t$  and has outstanding debt  $b_{t-1}$ ; income realization  $y_t$ ; expenditures  $e_t$ ; and faces the bond price menu  $q_t(b, y)$ . In addition, let  $V^d(b_{t-1}, y_t, e_t, q_t(b_t, y_t), y^*)$  be the value function of a government in default where  $b_{t-1}$  is the outstanding debt and  $y^*$  is the level of income at which the default decision was initially taken. The inclusion of  $y^*$  anticipates the fact that it partly determines the evolution of the bond price menu  $q_t(b_t, y_t)$  in equilibrium. Lastly, let  $d_t$  be an indicator variable that takes value 1 in case the country is in default at the end of time  $t$ .

Starting from the case where the sovereign is not in default in period  $t$ , the value functions have to satisfy:

$$V(b_{t-1}, y_t, e_t, q_t(b, y)) = \max_{(d_t, g_t, b_t) \in \mathbb{C}(b_{t-1}, y_t, e_t, q_t(b, y))} (1 - d_t) (u(c_t, g_t) + \beta E \{V(b_t, y_{t+1}, e_{t+1}, q_{t+1}(b, y))\}) \dots \\ + d_t (u(c_t, g_t) + \beta E \{V^d((1 + r)b_{t-1}, y_{t+1}, e_{t+1}, q_{t+1}(b, y), y_t)\}) \quad (2)$$

where

$$c_t = (1 - \tau)(y_t - d_t \alpha y_t)$$

and  $\mathbb{C}(b_{t-1}, y_t, e_t, q_t(b, y))$  is the feasibility set. It is fully characterized by the constraints:

$$\begin{aligned} g_t + e_t &\leq \tau(y_t - d_t \alpha y_t) + (1 - d_t)(q_t(b_t, y_t)b_t - b_{t-1}) \\ d_t &\in \{0, 1\} \end{aligned}$$

<sup>17</sup>Most sovereign debt models attribute to the government the economy-wide resource constraint, assuming that they can fully tax income with lump sum transfers. I generalize that setup and allow for the possibility of a government that has a limited ability to raise tax revenues and which expenditures might not be perfectly substitutable with private consumption. This approach allows me to have a more meaningful notion of debt sustainability through a lower flexibility of the government to adjust revenues and expenditures along the cycle. The standard framework can be recovered by setting  $\tau = 1$  and assuming  $c_t$  and  $g_t$  are perfect substitutes.

<sup>18</sup>This is a simplified version of the bargaining protocols presented in Yue (2010); Bai and Zhang (2012). It could be simply stated as a take-it-or-leave-it repayment offer with a lower bound. This is consistent with the view presented in Panizza, Sturzenegger, and Zettelmeyer (2009) where the authors claim that “ [after the Brady bonds] Debt restructurings took the form of a take-it-or-leave-it exchange offers, ...” (Panizza, Sturzenegger, and Zettelmeyer (2009), page 671).

Note that in case of default the face value of the outstanding debt is multiplied by the risk free interest rate  $(1 + r)$ . Therefore the debt takes into account the opportunity cost for creditors of the delay in any payment. I follow this procedure to have that any partial repayment of the outstanding debt correspond to a market value haircut and measures properly the losses incurred by creditors. Now, considering the case where the government starts the period in default. Then we have that the value functions satisfy:

$$V^d(b_{t-1}, y_t, e_t, q_t(b, y), y^*) = \max_{(d_t, g_t, b_t) \in \mathbb{C}^d(b_{t-1}, y_t, e_t, q_t(b, y))} (1 - d_t) [u(c_t, g_t) + \beta E \{V(b_t, y_{t+1}, e_{t+1}, q_{t+1}(b, y))\}] \dots \quad (3)$$

$$+ d_t [u(c_t, g_t) + \beta E \{V^d((1 + r)b_{t-1}, y_{t+1}, e_{t+1}, q_{t+1}(b, y), y^*)\}]$$

where

$$c_t = (1 - \tau)(y_t - d_t \alpha y_t)$$

and  $\mathbb{C}^d(b_{t-1}, y_t, e_t, q_t(b, y))$  is the feasibility set in default. It is given by

$$g_t + e_t + (1 - d_t) ((1 - \bar{h})b_{t-1} - q_t(b_t, y_t)b_t) \leq \tau(y_t - d_t \alpha y_t) \quad (4)$$

$$q_t(b_t, y_t)b_t \leq (1 - \bar{h})b_{t-1} \quad (5)$$

The constraint in equation (4) establishes that expenditures have to be lower or equal than tax revenues. Note that the repayment of the settlement can be decomposed as follows

$$(1 - \bar{h})b_{t-1} = q_t(b_t, y_t)b_t + ((1 - \bar{h})b_{t-1} - q_t(b_t, y_t)b_t)$$

where the first term on the right hand side is the amount of the repayment done with new bonds and the second term is the immediate payment. The second constraint (5) establishes that while in default, newly issued debt can only be used as part of the settlement.

Let  $\mathbf{s}_t = (b_{t-1}, y_t, e_t, q_t(b, y), y^*, d_{t-1})$  be the vector of states for the decision at  $t$ . We note the policy functions that solve the problem 3 of the government by  $d(\mathbf{s}_t)$ ,  $g(\mathbf{s}_t)$  and  $b(\mathbf{s}_t)$ .

Another definition that is going to prove useful is the following: Let  $\varrho_t$  be the market value of the bonds maturing at  $t$  once the default decision is announced. If the government decides to fully repay the bonds then the market value is equal to the face value and  $\varrho_t = 1$ . Otherwise, if the government announces a default then  $\varrho_t = 1 - \bar{h}$ . Irrespectively of when the settlement is achieved, the creditors know a haircut  $\bar{h}$  over the market value is going to be imposed and consequently  $\varrho_t = 1 - \bar{h}$ .

### Perceived Law of Motion of the Bond Price Menu

To solution of (2) and (3) requires to specify how the government forms expectations over  $y_{t+1}$ ,  $e_{t+1}$  and  $q_{t+1}(b, y)$ . I suppose the government knows the stochastic processes of  $y_t$  and  $e_t$ . On the other hand the expectations over the dynamics of the bond price menu are given by the perceived law of motion:

$$q_{t+1}(b, y) = L(b_{t-1}, y_t, e_t, q_t(b, y), d_t, d_{t-1}, y^*)$$

The selection of the arguments of the function  $L$  seems arbitrary so far. It is written in a general way such that given the learning mechanism of the creditors it comprises rational expectations of the government and also allows for potential deviations from it, e.g a misspecified model. To further discuss the nature of  $L$  I introduce next the other agents in this economy: the creditors.

### 3.2 The Creditors and the equilibrium bond price menu

International creditors are risk neutral, competitive and have access to a risk free asset with gross return  $1 + r$ .

The system of beliefs of the creditor is given by a probability measure  $\mathcal{P}$  that specifies the joint distribution of  $\{\varrho_t, y_t, b_t\}_{t=0}^{\infty}$ .  $\mathcal{P}$  is given and a primitive of the analysis and does not necessarily coincides with the distribution of  $\{\varrho_t, y_t, b_t\}_{t=0}^{\infty}$  in equilibrium. Taking  $\mathcal{P}$  as given the creditors maximize their utility, following the concept of Internal Rationality (Adam and Marcet (2011)).

Because of risk neutrality investors only take into account the expected return of the bond for their trading decision. The creditors' expected market value of the bonds  $b_t$  issued when the country income is  $y_t$  corresponds to  $E_t^{\mathcal{P}}\{\varrho_{t+1}\}$ , where  $E_t^{\mathcal{P}}$  is the expectation operator with the probability measure  $\mathcal{P}$  and conditional on the history  $\{\varrho_s, y_s, b_s\}_{s=0}^t$ . The expected market value can be written as:

$$E_t^{\mathcal{P}}\{\varrho_{t+1}\} = \sum_{y \in Y} E_t^{\mathcal{P}}\{\varrho_{t+1} | y_{t+1} = y\} P_t^{\mathcal{P}}\{y_{t+1} = y\} \quad (6)$$

where  $P_t^{\mathcal{P}}\{A\}$  is the probability of event  $A$  with the probability measure  $\mathcal{P}$ . Consequently the expected return of the bond corresponds to

$$\frac{E_t^{\mathcal{P}}\{\varrho_{t+1}\}}{q_t(b_t, y_t)} = \frac{\sum_{y \in Y} E_t^{\mathcal{P}}\{\varrho_{t+1} | y_{t+1} = y\} P_t^{\mathcal{P}}\{y_{t+1} = y\}}{q_t(b_t, y_t)}$$

and assuming all investors have the same system of beliefs  $\mathcal{P}$  we have that the non-arbitrage condition is:

$$q_t(b_t, y_t) = \frac{\sum_{y \in Y} E_t^{\mathcal{P}}\{\varrho_{t+1} | y_{t+1} = y\} P_t^{\mathcal{P}}\{y_{t+1} = y\}}{1 + r}$$

this equation characterizes the equilibrium bond price menu  $q_t(b_t, y_t)$  at time  $t$ . The time varying nature of  $q_t(b, y)$  is inherited from  $E_t^{\mathcal{P}}\{\varrho_{t+1} | y_{t+1} = y\}$  and  $P_t^{\mathcal{P}}\{y_{t+1} = y\}$ . Next we show what are the constraints that the REE imposes on  $\mathcal{P}$  and the deviation we propose to study sovereign debt markets.

### Recursive Rational Expectations Equilibrium (REE) spreads

First I describe the “Objective” expected value of repayment implied by the model. Let  $\mu_t(\mathbf{s}_{t-1}, y_t)$  be the mean of repayment at  $t$  conditional on  $t - 1$  states  $\mathbf{s}_{t-1}$  and contemporaneous income  $y_t$ . Then, given government policy functions, the market value of bonds is

$$E_t \{\varrho_{t+1}\} = \sum_{y \in Y} \mu_{t+1}(\mathbf{s}_t, y) \pi \{y \mid y_t\}$$

where

$$\mu_{t+1}(\mathbf{s}_t, y) = \sum_{e \in E} (1 - \bar{h} d_{t+1}(\mathbf{s}_{t+1})) f(e)$$

where we are averaging over all the possible realizations of the expenditures shock  $e$ , with its corresponding probability mass function  $f(e)$ . Note that the previous system of equations average over the two only exogenous stochastic processes in the model: income  $y$  and the expenditures  $e$ .

Now I can proceed to define the Recursive REE with a constant bond price menu.<sup>19</sup>

**Recursive-REE** A set of government policy functions for *i*) default  $d(\mathbf{s})$ ; *ii*) public goods provision  $g(\mathbf{s})$ ; *iii*) debt issuance  $b(\mathbf{s})$ ; and *iv*) a bond price menu  $q(b, y)$  such that:

1. Taking as given the bond price menu, the government policy functions  $d(\mathbf{s})$ ,  $g(\mathbf{s})$ ,  $b(\mathbf{s})$  and  $h(\mathbf{s})$  solve the government’s problem for every  $t$ .
2. The bond price menu  $q(b, y)$  satisfies the following condition

$$q(b, y) = \frac{\sum_{y' \in Y} \mu(b, y') \pi \{y' \mid y\}}{1 + r}$$

where

$$\mu(b, y) = \sum_{e \in E} (1 - \bar{h} d(\mathbf{s}_{t+1})) f(e)$$

3. The probability measure  $\mathcal{P}$  is equal to the distribution of  $\{\varrho_t, y_t, b_t\}_{t=0}^{\infty}$  implied by the policy functions and the process for the exogenous variables  $y_t$  and  $e_t$ .

The first condition refers to individual rationality given beliefs. The second condition guarantees that creditors break even and the bond price menu reflects the repayment probabilities. Condition 3 establishes that the creditors beliefs are rational expectations.

Condition 3 can be relaxed and still sustain the same allocation. Note that the creditor decision making depend on two sufficient statistics at any given period  $E_t^{\mathcal{P}} \{\varrho_{t+1} \mid y_{t+1} = y\}$  and  $P_t^{\mathcal{P}} \{y_{t+1} = y\}$ . Then the requirements on  $\mathcal{P}$  to support the rational expectations allocation are

<sup>19</sup>This definition closely follows the equilibrium concept prevalent in this literature (see Arellano (2008))

the following:

$$\begin{aligned} E_t^{\mathcal{P}} \{\varrho_{t+1}\} &= \mu(b_t, y) \\ P_t^{\mathcal{P}} \{y_{t+1} = y\} &= \pi \{y' \mid y\} \end{aligned}$$

these constraints establish that subjective expected repayment has to coincide with average repayment and the perceived transition probabilities of income coincide with the objective transition probabilities. Next we explore a deviation of the first requirement and formulate a setup where creditors learn about the repayment probabilities.

### Subjective beliefs

The system of subjective beliefs is characterized by the following:

1. Creditors beliefs that the process for the market price of maturing bonds when outstanding debt is  $b$  at the level of GDP  $y'$  corresponds to:

$$\begin{aligned} \varrho_t(b, y') &= m_t(b, y') + \epsilon_t \\ m_t(b, y') &= m_{t-1}(b, y') + u_t \end{aligned} \tag{7}$$

where  $\epsilon_t \sim N(0, \sigma_\epsilon)$  and  $u_t \sim N(0, \sigma_u)$  are independent from each other and also from  $\{\varrho_t, y_t, b_t\}_{s=0}^t$ .

2. The subjective beliefs coincide with the objective transition probabilities

$$P_t^{\mathcal{P}} \{y_{t+1} = y\} = \pi \{y' \mid y\}$$

The shocks  $\epsilon_t$  and  $u_t$  are not specific to the contingency  $(b, y')$  and consequently the creditors extract information from a default situation about the whole bond price menu. In the appendix A I consider the possibility of specific shocks for each contingency.<sup>20</sup>

The Recursive-REE can be recovered by imposing that  $m_0(b, y') = \mu(b, y')$  and setting the variance of the shocks  $u_t$  to zero. The deviation I consider from Rational Expectations is to allow  $\sigma_u > 0$  and equal to a small fraction of the variance of  $\epsilon_t$ .

According to 7 equation 6 corresponds to

$$E_t^{\mathcal{P}} \{\varrho_{t+1}\} = \sum_{y \in Y} \hat{m}_t(b_t, y) \pi \{y' \mid y\}$$

<sup>20</sup>This specification 7 is also found in Adam, Marcet, and Nicolini (2016) that propose it for the price dynamics of the stock. Notice the difference here that it is specified for the whole bond price menu, i.e for all possible combinations  $(b, y')$ .

where  $\hat{m}_t(b, y)$  is the mean of the posterior beliefs of the creditor about  $m_t(b, y)$ . I set the prior of the creditors to be a degenerate distribution on the REE bond price menu at period 0. The optimal updating of beliefs is the following fixed gain learning algorithm,

**Learning Algorithm:**

$$\hat{m}_t(b, y) = \begin{cases} \hat{m}_{t-1}(b, y) + \kappa (\varrho_t - \hat{m}_{t-1}(b_{t-1}, y_t)) & \text{if } d_{t-1} = 0 \\ \hat{m}_{t-1}(b, y) & \text{else} \end{cases} \quad (8)$$

where  $\kappa = \frac{\sigma_\epsilon}{\sigma_\epsilon + \sigma_u}$ . The first case refers to the periods where bonds debt matures and the second for when the country is in default. Consequently the bond price menu is updated as follows

$$q_t(b, y) = \begin{cases} q_{t-1}(b, y) - \frac{\kappa}{1+r} (\varrho_t - \mu_{t-1}(b_{t-1}, y_t)) & \text{if } d_{t-1} = 0 \\ q_{t-1}(b, y) & \text{else} \end{cases}$$

The question that emerges at this point is if the government perceived dynamics of the bond price menu law of motion of the bond price menu, given by  $q_{t+1} = L^g(b_{t-1}, y^*, e_t, q_t, d_t, d_{t-1})$  coincides with the actual updating process described in the previous equations. If the perceived dynamics do coincide with the updating process then we have that reputational concerns of the government will affect the debt issuance and repayment policies. We will study this in section 6 to evaluate if it provides the incentives of an insurance contract and in section 7 to derive an empirical test for it. In the next section this mechanism is not main driver of the points highlighted and I abstract from it by having sovereigns that do not internalize the effects of their actions on the bond price menu or equivalently having a perceived law of motion given by  $L(b_{t-1}, y^*, e_t, q_t, d_t, 1) = q_t$ .

## 4 Spreads after default

In this section I will illustrate the pattern of spreads implied by the bond price menu  $q(b, y)$  by simulations of the model for specific functional forms and parameter values. The parametrization is presented in table 1. The model is solved by value function iterations for each element of the sequence of the bond price menu  $q_t(b, y)$ . One key assumption at this stage is that governments have point expectations about the bond price menu in the future that are equal to the observed bond price menu.

Figure 1 show the time series of one of the entries of the bond price menu. The price shown is for the bond issued with a high level of debt and low  $y$ . The fixed point equilibrium is presented too for comparison.

The bond price fluctuate around the fixed point equilibrium but the deviations can be persistent. The “mean-reversion” property is endogenous in the model and generated by the optimal debt and default decisions of the government. This element is easier to discuss if we think in terms of the risk spreads (see figure 2) implied by the bond price. When spreads are low the government is willing

Table 1: Parameter Values and functional forms

$\alpha$	0.9
$\beta$	0.95
$\tau$	0.2
$y$	$[1, 2.5]$
$P(y'   y)$	$\begin{pmatrix} 0.3 & 0.1 \\ 0.7 & 0.9 \end{pmatrix}$
$b$	$[0.1, 0.2, 0.4]$
$e$	$[-0.5, 0.17]$
$P(e)$	$[0.7, 0.3]$
$h$	0.5
$r$	0.03
$\kappa$	0.2
$\gamma$	0.01
$u(c, g)$	$\left(c^{\frac{\theta-1}{\theta}} + g^{\frac{\theta-1}{\theta}}\right)^{\frac{\theta}{\theta-1}\theta_2}$
$\theta$	0.7
$\theta_2$	0.5

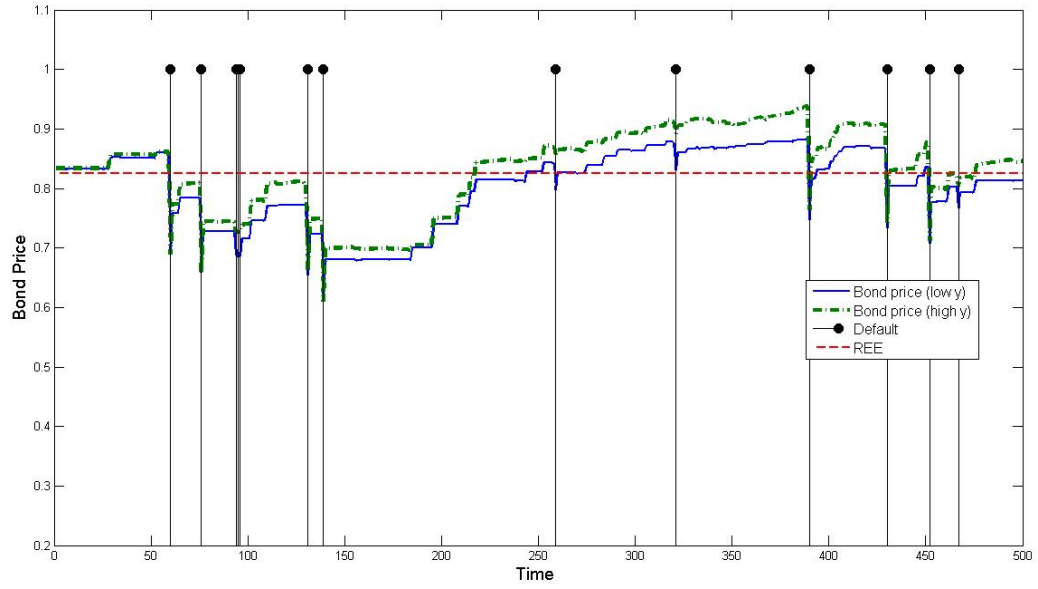
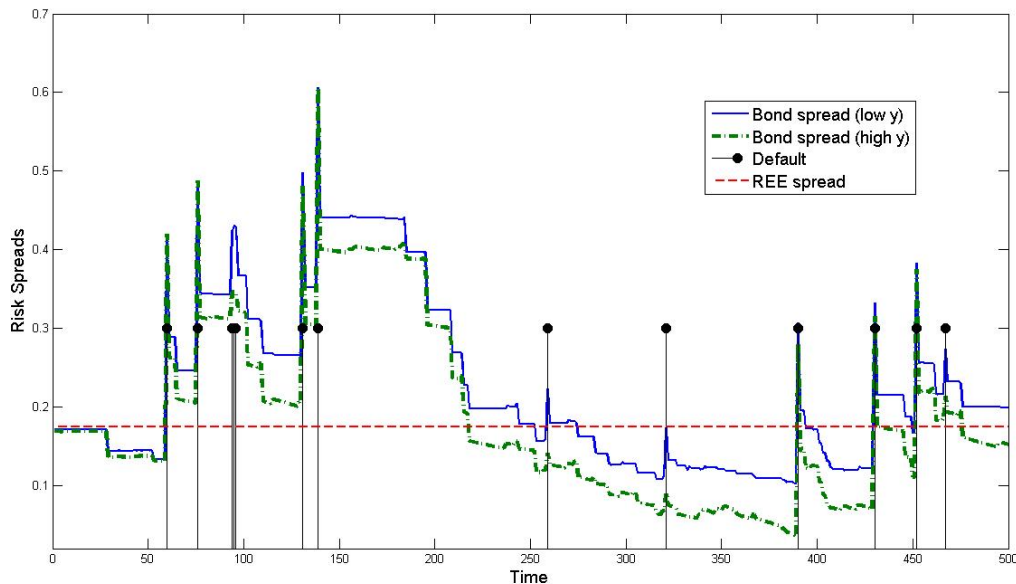
Figure 1: Bond prices and default  $q(b, y)$ 

Figure 2: Risk spreads and default



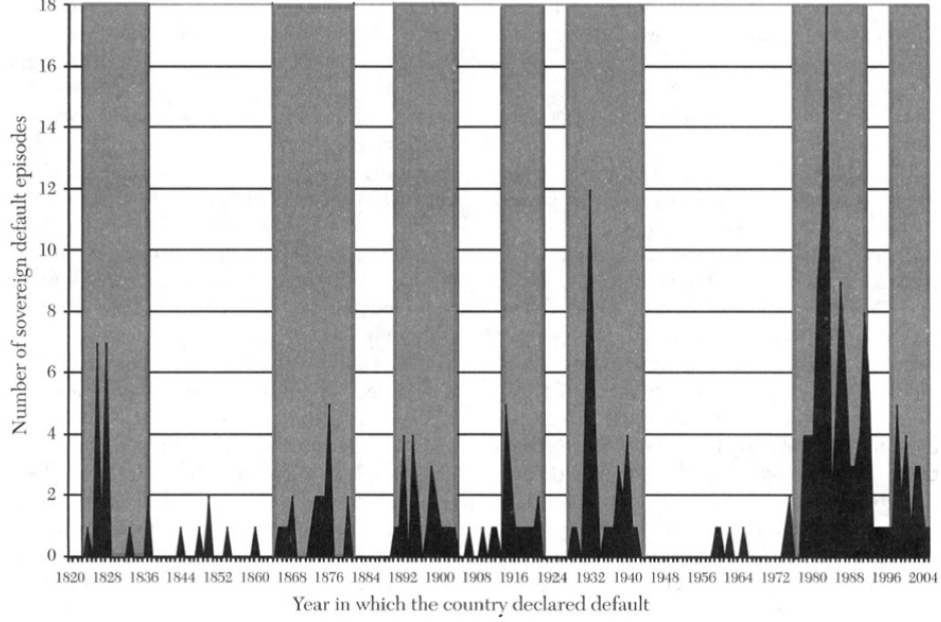
to issue large amounts of debt and be more exposed to an expenditures shock that might lead to default because the increase in the current level of resources is high enough to compensate for the additional risk. On the other hand, for high spreads the country default become less likely, despite the reduced ability to roll over debt, because the government decides to “gamble” less. In other words, default is the correction mechanism when spreads are low and repayment when spreads are high.

We can observe how after a sequence of defaults spreads rise and can lead to a surplus for the creditors by reacting beyond the objective default probability. In a REE spreads would be constant at the presented level reflecting exactly the default probability of the sovereign for that bond price menu.

One of the challenges of the sovereign debt literature is to construct models that are able to generate realistic levels of debt to GDP. The models tend to support only low levels of debt compared to what we observe in reality. In our case, when the bond prices are close to the REE the debt levels tend to be low. It is when the economy has lower spreads that the REE spreads that high debt and the subsequent default emerge naturally. These cycles of sovereign spreads opens the possibility that if those cycles are somehow coordinate among a group of countries then clusters of default can emerge as is explored in the next section.



Figure 3: Clusters of default



Taken from: Panizza, Sturzenegger, and Zettelmeyer (2009)

## 5 Clusters of default

Now that we have a model that captures the movements in spreads after default, we can move to study other dimensions of the sovereign debt market and evaluate what can we extract from the proposed mechanism. In this section I use the model to study one of the main features of this market: defaults tend to happen in clusters, and as mentioned in Panizza, Sturzenegger, and Zettelmeyer (2009) “typically following the end of a period of rapid credit expansion to the borrowing countries”. figure 3 presents the frequency of defaults in the last two centuries.

To study the possibility of clusters of default with our model we extend the analysis to a pool of five equal countries. The system of beliefs of the creditor is modified to allow for the possibility of a common component in the default risk of this group of countries. Creditors system of beliefs for the repayment decision of country  $i$  in period  $t$  is characterized by

$$\phi_t^i(b, y') = \mu_t^i(b, y') + \Omega_t(b, y') + \epsilon_t^i \quad (9)$$

$$\mu_t^i(b, y') = \mu_{t-1}^i(b, y') + u_t^i$$

$$\Omega_t(b, y') = \Omega_t(b, y') + \nu_t \quad (10)$$

where  $\Omega_t(b, y')$  is a common component for all the countries  $i \in \{1, \dots, 5\}$  and  $\mu_t^i(b, y')$  is the idiosyncratic components. The shocks  $\epsilon_t^i, u_t^i$  are idiosyncratic and independent across countries. Therefore the only common shock is  $\nu_t$ .

As in the original model the REE can be recovered as a particular case of these setup. By having the variance of the shocks  $u_t^i$  and  $\nu_t$  to be zero and

$$\mu(b, y') = \mu_t^i(b, y') + \Omega_t(b, y')$$

such that the bond menu is the same for all countries, does not change across time and is equal to the REE value.

In this case the creditors update their perceived default probabilities of country  $i$  taking into account not only the repayment outcome of country  $i$  but potentially also the repayment outcomes in the other countries. The optimal updating algorithm in this scenario is the following:

$$\mu_t^i(b, y) = \mu_{t-1}^i(b, y) + \kappa^i (R_t^i(b_{t-1}^i, y_t^i) - \mu_{t-1}^i(b_{t-1}^i, y_t^i)) + \kappa^o \sum_{j \neq i} (R_t^j(b_{t-1}^j, y_t^j) - \mu_{t-1}^j(b_{t-1}^j, y_t^j)) \quad (11)$$

where  $\kappa^i$  is the gain factor for repayment about the same country and  $\kappa^o$  for all the other sovereigns. These gain factor satisfy  $\frac{\kappa^o}{\kappa^i} \sim \frac{\sigma(u_t^i)}{\sigma(\nu_t)}$ , the reaction of spreads to information from another country depends on how large the creditor perceive common factors are responsible for the variance of the country outcomes.

First consider the case where the variance of  $\nu_t$  is zero. In this scenario repayment of country  $j$  is not informative about country  $i$ . In this case creditors learn about each country independently. The figure 4 shows the number of defaults in a rolling window of 10 periods, and the average risk spread for high debt. In this case clusters can happen just by chance and we see that there is only one episode where 4 countries defaulted.

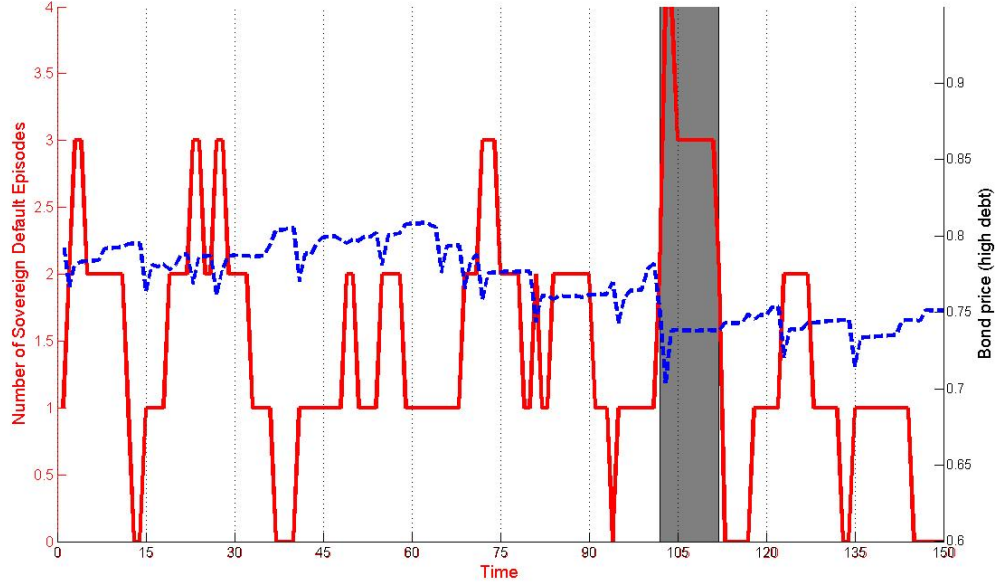
## Cross country learning

Now we consider the case where the common component has positive variance and therefore creditors perceive there is a common factor that drives countries default decision. I maintain the assumption that the idiosyncratic shock is more volatile than the aggregate shock and therefore  $\kappa^i > \kappa^o$ .

The figure 5 show the number of defaults in a rolling window of 10 periods and the average risk spread for high debt. Clusters occur now more frequently and they tend to be preceded by periods of increasing bond prices for high debt and low levels of defaults.

The default clusters are not simply driven by a sudden burst of pessimism (in creditors) and contagion after one country defaults. In fact, what leads them to default contemporaneously is the coordination in high debt exposure. Note that when creditors observe few or no defaults (like the first part of the sample) over a long period of time they tend to become too “optimistic”. This can be seen in the positive trend of the bond prices, that leads to an increase of all governments willingness to issue more debt. The increase in debt increases their exposure and when a bad shocks occurs

Figure 4: Clusters without cross-country learning



the government defaults. So, the arrival of bad shocks and the effect of the subsequent defaults in the cost of roll over the debt is the final element that generates the cluster.

The previous graphs focus on the relation between the bond price menu and defaults or clusters of defaults, and the highlighted mechanism is that when the entries of bond price menu increase (specially those associated with high debt) the governments decide to issue more debt and are more exposed to a negative expenditures shock or a fall in GDP. Empirical evidence has shown that in fact spreads rise before defaults as the countries face more risk and/or issue more debt, this is coherent with the mechanism proposed here because the empirical evidence points out to realized spreads rather to the bond price menu. figure 6 shows how the realized spreads do increase before the cluster of defaults and that effectively the mechanism proposed is driving the results as we can see the high debt exposure that countries face before the clusters of defaults. What this paper is pointing out is something that cannot be extracted that easily from the data and that is the underlined dynamics of the bond price menu rather than those of the realized bond prices, where the literature so far has focused assuming a fixed and constant bond price menu.

## 6 Discussion of Results, caveats and alternative theories

The first of the two stylized facts discussed (the raise in spreads after default) is a pattern that has been extensively studied and for which Cruces and Trebesch (2013) provide precise evidence for the period 1970-2010. On the other hand the second stylized fact has not brought as much attention

Figure 5: Clusters with cross-country learning

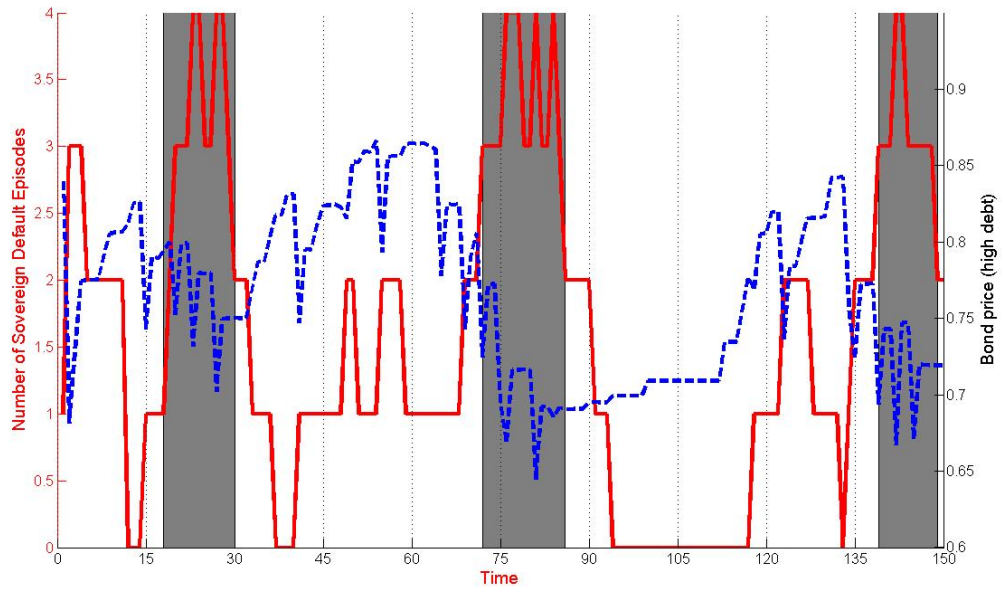
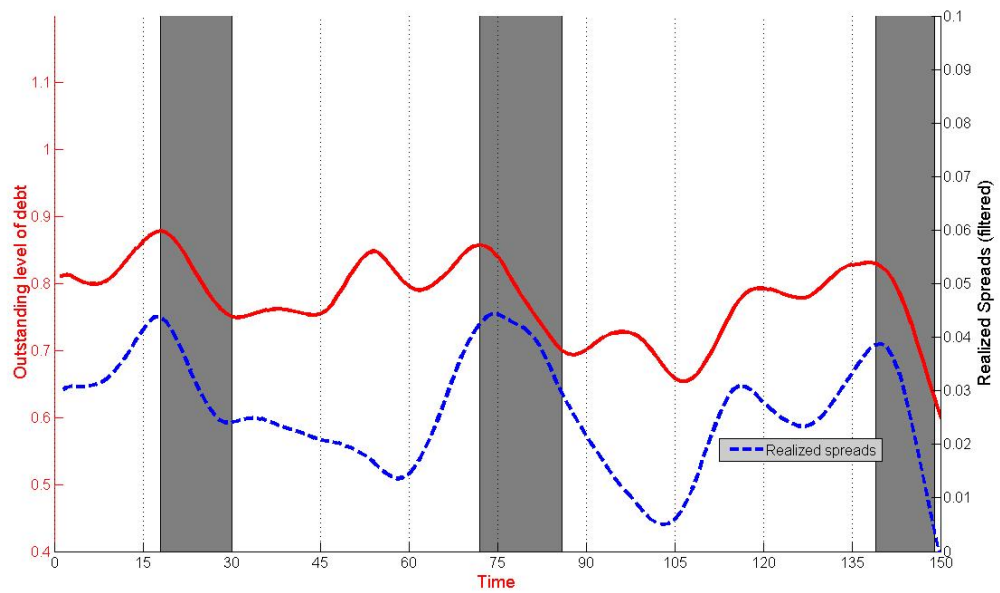


Figure 6: Total debt issued and the realized spreads (HP filtered)



and Alfaro and Kanczuk (2005) find it with a limited sample 1970-1982 where most of the lending was done by syndicated bank loans. An extension of their analysis with the sample used by Cruces and Trebesch (2013) would allow to identify if the results are robust to periods where most of the lending was done using sovereign bonds, as the theory I propose suggests, or if it is something particular of syndicated bank loans and an outcome of relational contracts as Alfaro and Kanczuk (2005) suggests. The fact that despite the different methodologies and samples in both studies the authors find very similar responses of spreads to default could indicate that indeed it is a robust finding.

The paper provides a theory that is based on a belief system of the creditors in sovereign debt markets. The model ability to replicate the stylized facts proposed is shown with simulations, but a better understanding of the mechanism and its interaction with the government behavior demands a more complete characterization of equilibrium outcomes.

Compared to Adam, Marcet, and Nicolini (2016) the use of the Internal Rationality concept in this setup is extremely simplified by the assumptions of risk neutrality and the implied “deep pockets” creditors. These conditions render the problem of the creditors basically static and it is only beliefs that are carried from one period to the other. Furthermore, uncertainty about the repayment plays no role in the creditors decisions and what matters is only the expected repayment. It is an open question if the model can generate quantitatively reasonable levels of spreads as it is.

An evaluation to assess if creditors’ beliefs system are not rejected by the model’s simulated data or the empirical behavior of spreads is a necessary test to assess the consistency and empirical relevance of the modelling procedure. The fact that the bond price menu is potentially changing every period could in principle sustain a process for default probabilities that could resemble the unit root imposed on the belief system, although a formal test in this dimension could shed light on the validity of the belief system.

The default clusters that have appeared historically have also coincided in many cases with large common shocks to the defaulting sovereigns. A test of the existence of such channel as cross country learning nevertheless could explain the excess comovement in the spreads of emerging markets that is not accounted for global factors; as local conditions have shown to be a weak explanatory variable in this dimension because of the weak correlation of output.

Extensions of the analysis that are ongoing work are: i) whether the implied learning algorithm implements an allocation that resembles the optimal insurance contract with asymmetric information. ii) The analysis of multi-period bonds and test whether sovereign’s reputation considerations hold in the data.

## 7 Conclusions

Creditors in international markets price sovereign debt contracts with great uncertainty about the conditions under which repayment is going to be done, and about the likelihood of each of those

conditions. The sources of uncertainty range from the structure of the economy to the political decision making process. This complicated and demanding endeavor of pricing sovereign debt has a simple and direct source of information, the past behavior of the country and similar countries. Which debt contracts were paid and which others were not is a direct measure of the relevant statistic that concerns creditors: the default probability.

With a model where creditors learn about the default probabilities of debt contracts directly from the past behavior of a country, I showed that the pattern of spreads after default can be the outcome of the creditors' updating of beliefs. This channel can be seen as the reputation cost of the default decision.

The model replicates the fact that spreads tend to increase after defaults and also the possible emergence of a surplus for creditors. The mechanism proposed implies a certain degree of symmetry that prescribes that after long periods where creditors have experienced full repayment spreads can fall and generate a negative surplus for creditors. This provides a rationale for the view that debt clusters come as a result of periods of credit expansions where risk spreads do not reflect the default probability and the high debt positions of countries.

## References

- ADAM, K., AND A. MARCET (2011): “Internal rationality, imperfect market knowledge and asset prices,” *Journal of Economic Theory*, 146(3), 1224–1252.
- ADAM, K., A. MARCET, AND J. P. NICOLINI (2016): “Stock Market Volatility and Learning,” *The Journal of Finance*, LXXI(1).
- AFONSO, A., P. GOMES, AND P. ROTHER (2006): “What Hides Behind Sovereign Debt Ratings?,” Working Papers 2006/35, Department of Economics at the School of Economics and Management (ISEG), Technical University of Lisbon.
- AGUIAR, M., AND M. AMADOR (2014): “Chapter 11 - Sovereign Debt,” in *Handbook of International Economics*, ed. by E. H. Gita Gopinath, and K. Rogoff, vol. 4 of *Handbook of International Economics*, pp. 647 – 687. Elsevier.
- AGUIAR, M., AND G. GOPINATH (2006): “Defaultable debt, interest rates and the current account,” *Journal of International Economics*, 69(1), 64–83.
- ALFARO, L., AND F. KANCZUK (2005): “Sovereign debt as a contingent claim: a quantitative approach,” *Journal of International Economics*, 65(2), 297–314.
- ARELLANO, C. (2008): “Default Risk and Income Fluctuations in Emerging Economies,” *American Economic Review*, 98:3, 690–712.
- BAI, Y., AND J. ZHANG (2012): “Duration of sovereign debt renegotiation,” *Journal of International Economics*, 86:2, 252–268.
- BENCZÚR, P., AND C. L. ILUT (2016): “Evidence For Relational Contracts In Sovereign Bank Lending,” *Journal of the European Economic Association*, 14(2), 375–404.
- BORENSZTEIN, E., AND U. PANIZZA (2009): “The Costs of Sovereign Default,” *IMF Staff Papers*, 56(4), 683–741.
- CANTOR, R., AND F. PACKER (1996): “Determinants and impacts of sovereign credit ratings,” Research Paper 9608, Federal Reserve Bank of New York.
- CATÃO, L. A., A. FOSTEL, AND S. KAPUR (2009): “Persistent gaps and default traps,” *Journal of Development Economics*, 89(2), 271–284.
- CRUCES, J. J., AND C. TREBESCH (2013): “Sovereign Defaults: The Price of Haircuts,” *American Economic Journal: Macroeconomics*, 5(3), 85–117.
- EVANS, G., AND S. HONKAPOHJA (2001): *Learning and Expectations in Macroeconomics*. Princeton University Press.
- FLANDREAU, M., AND F. ZUMER (2003): “The Making of Global Finance 1880-1913,” Discussion paper, OECD Development Center Research.

- GROSSMAN, H. I., AND J. B. VAN HUYCK (1988): “Sovereign Debt as a Contingent Claim: Excusable Default, Repudiation, and Reputation,” *American Economic Review*, 78(5), 1088–97.
- KLETZER, K. M., AND B. D. WRIGHT (2000): “Sovereign Debt as Intertemporal Barter,” *American Economic Review*, 90(3), 621–639.
- PANIZZA, U., F. STURZENEGGER, AND J. ZETTELMEYER (2009): “The Economics and Law of Sovereign Debt and Default,” *Journal of Economic Literature*, 47(3), 651–98.
- RICHMOND, C., AND D. A. DIAS (2008): “Duration of Capital Market Exclusion: Stylized Facts and Determining Factors,” Discussion paper, Inter-American Development Bank.
- TOMZ, M. (2007): *Reputation and International Cooperation: Sovereign Debt Across Three Centuries*. Princeton University Press.
- TOMZ, M., AND M. L. WRIGHT (2013): “Empirical Research on Sovereign Debt and Default,” *Annual Review of Economics*, 5(1), 247–272.
- TOMZ, M., AND M. L. J. WRIGHT (2007): “Do Countries Default in “Bad Times” ?,” *Journal of the European Economic Association*, 5(2-3), 352–360.
- WRIGHT, M. L. J. (2005): “Coordinating Creditors,” *American Economic Review*, 95(2), 388–392.
- YUE, V. Z. (2010): “Sovereign default and debt renegotiation,” *Journal of International Economics*, 80(2), 176–187.



## A System of beliefs with uncorrelated shocks

### Subjective beliefs

The system of subjective beliefs is given by the following equations: For a given outstanding debt  $b$  at the level of GDP  $y'$  the creditor beliefs that the haircut in that situation is given by:

$$\begin{aligned} R_t(b, y') &= 1 - \mu_t(b, y') + \epsilon_t \\ \mu_t(b, y') &= \mu_{t-1}(b, y') + u_t \end{aligned} \tag{12}$$

for each possible debt level  $b \in B$  and GDP level  $y' \in Y$ . Where  $\epsilon$  and  $u_t$  have mean zero. Therefore, the creditors assign a random walk for the average payoff of each possible repayment situation characterized by a level of GDP  $y'$  and debt  $b$ . Note that this belief system collapses to the Rational expectations case if we set  $\theta_t(b, y') = \mu(b, y')$  and the variance of the shocks  $u_t$  is zero.

This specification leads to a joint distribution of the next period GDP  $y'$  and the expected repayment for each contract  $E_t \{R_{t+1}(b, y')\}$ . If the variance of shock  $u_t$  is set to zero then the expected repayment for each contract has a degenerate distribution that assign all the mass to the point  $\mu(b, y')$ .

From the beliefs system 12 it follows that the relevant information for the creditor to update his beliefs is the repayment outcomes  $R(b, y)$  for the corresponding contingency  $(b, y)$ . To be specific about the information contained in  $R(b, y)$  I characterize this mapping, when the country is not in default, as follows:

- $R_t(b_{t-1}, y_t) = 1$  in case of full repayment.
- $R_t(b_{t-1}, y_t) = 0$  if the government decide to default at  $t$ .

On the other hand, if the country starts the period in default and the settlement is achieved with the haircut level  $h$  we have

- $R_{t+s}(b_{t-1}, y_t) = 1 - h$ , where  $t$  is the period the country entered in default and  $s$  the number of periods before the settlement was achieved.

Given their beliefs system 12 and the available information creditors optimally update their beliefs with a fixed gain algorithm as follows:

At period  $t$  the debt level  $b_{t-1}$  matures; the realization of GDP is given by  $y_t$ ; and the repayment decision of the government is  $R_t(b_{t-1}, y_t)$ . Then the expected haircut for each contingency  $(b, y)$  is evolve as

$$\mu_t(b, y) = \begin{cases} \mu_{t-1}(b, y) + \kappa (R_t(b_{t-1}, y_t) - \mu_{t-1}(b_{t-1}, y_t)) & \text{if } (b, y') = (b_{t-1}, y_t) \quad \wedge \quad d_{t-1} = 0 \\ \mu_{t-1}(b, y) - \kappa R_t(b_{t-1}, y_t) & \text{if } (b, y') = (b_{t-1}, y_t) \quad \wedge \quad d_{t-1} = 1 \\ \mu_{t-1}(b, y') & \text{else} \end{cases} \quad (13)$$

the first two cases refer to the periods where an observation of repayment related to the debt contract  $(b, y)$  was realized, and the last case to when no new information about this contract arrived. The first case is the learning rule when the country is not in default ( $d_{t-1} = 0$ ) and  $\kappa$  is the gain factor. The second case incorporates the new information that comes from the settlement, and is multiplied by  $-\kappa$  to avoid "double counting" the default episode, if we consider the whole default episode as a single event the updating rule is equivalent to  $\mu_{t+s}(b, y) = \mu_t(b, y) + \kappa (h - \mu_t(b_{t-1}, y_t))$ , a standard fixed gain rule.

The fact that  $\mu(b, y)$  is updated twice during a default episode, i.e. in the period default is chosen and in the period the settlement is achieved, instead of only once (say, at the end of the default episode) implies the distinctive feature that bonds that are included in the settlement are valued less than the same type of bond (equal  $y$  and  $b$ ) after the settlement is achieved and the country return to international debt markets.

There is one shortcoming of the updating rule (13), every period almost all the contracts will lie in the third case and will not be updated; only the observed contract repaid is updated. The learning approach in this paper seeks to stress the idea that creditors might not know perfectly the problem the government solves to decide repayment, nevertheless the fact that the creditor do not use the outcome in one particular contract to learn about other, possibly similar, contracts seems to strong. Therefore I propose a simple and intuitive inference rule for the creditor that allow him to link contracts, in particular to link the expected haircut for different levels of outstanding debt and GDP,

**Rule for inference:** *If  $b \geq \tilde{b}$  and  $y \leq \tilde{y}$  then  $\mu(b, y) \geq \mu(\tilde{b}, \tilde{y})$ .*

This rule models the idea that more debt and lower GDP are associated with higher risk of default. I incorporate this rule in the learning mechanism of the creditors using a two steps procedure where the first step is just the same as (13) and the second step adjust the beliefs of the contingencies that were not updated in the first step such that the rule for inference is satisfied. The procedure is described formally as follows:

Taking as given the beliefs at  $t - 1$  as  $\mu_{t-1}$  the steps for updating once the information at time  $t$  is revealed is:

1. Update the prior using the repayment decision

$$\hat{\mu}_t(b, y) = \begin{cases} \mu_{t-1}(b, y) + \kappa (R_t(b, y) - \mu_t(b, y)) & \text{if } (b, y') = (b_{t-1}, y_t) \quad \wedge \quad d_{t-1} = 0 \\ \mu_{t-1}(b, y) - \kappa R_t(b, y) & \text{if } (b, y') = (b_{t-1}, y_t) \quad \wedge \quad d_{t-1} = 1 \\ \mu_{t-1}(b, y') & \text{else} \end{cases}$$

2. Update the beliefs for such contingencies such that  $\hat{\mu}_t(b, y) = \mu_{t-1}(b, y')$  to satisfy the rule for inference

$$\mu_t(b, y) = \begin{cases} \max \left\{ \hat{\mu}_t(b, y), \hat{\mu}_t(\tilde{b}, \tilde{y}) \right\} & \text{if } b \geq \tilde{b} \quad \wedge \quad y \leq \tilde{y} \\ \min \left\{ \hat{\mu}_t(b, y), \hat{\mu}_t(\tilde{b}, \tilde{y}) \right\} & \text{if } b < \tilde{b} \quad \wedge \quad y \geq \tilde{y} \\ \hat{\mu}_t(b, y) & \text{else} \end{cases}$$

where  $(\tilde{b}, \tilde{y})$  is the duple of debt and GDP such that  $\hat{\mu}_t(b, y) = \mu_{t-1}(b, y')$  (the updated contingency).

This procedure allows to keep the updating rule (13) and satisfy the rule for inference with the minimum possible change in beliefs.

## A.1 Perceived law of motion $L$

I suppose governments have rational expectations about  $q_t$  and therefore the perceived law of motion of  $q_t$  coincides with the objective or "true" law of motion.

Recall that  $L$  is a mapping from the space of the states  $(b_{t-1}, y_t, e_t, q_t, d_t, d_{t-1})$  to the space  $B \times Y \times \mathbb{R}_+$ . I define  $L^{(b, y)}$  as the image of  $L$  with the two first entries fixed at  $(b, y)$  such that we have  $q_{t+1}(b, y) = L^{(b, y)}(b_{t-1}, y_t, e_t, q_t, d_t, d_{t-1})$ . Using the setup of the creditor the law of motion of  $q_t(b, y)$  can be characterized as follows (check the appendix for details):

- Case 1: The government starts the period  $t$  not in default i.e  $d_{t-1} = 0$

$$\begin{aligned} q_t(b, y) &= L(b_{t-1}, y_t, e_t, q_t, d_t, d_{t-1}(=0)) \\ &= q_{t-1}(b, y) + \kappa \Gamma_t(b_{t-1}, y_t, e_t, q_t, d_t, d_{t-1}(=0)) - \kappa(d_t) \frac{\pi_t(y_t | y)}{1+r} \end{aligned}$$

where  $\Gamma \geq 0$  and measure the gain in bond prices for repayment. For example if only two realizations of  $y$  can occur it corresponds to:

$$\Gamma = \frac{\pi_t(y^* | y)}{1+r} \frac{(1+r) \left[ q_t(b^*, y') - q_t(b^*, y^*) \frac{\pi_t(y' | y')}{\pi_t(y' | y^*)} \right] - \left( 1 - \frac{\pi_t(y' | y')}{\pi_t(y' | y^*)} \right)}{\left[ \frac{\pi_t(y' | y')}{\pi_t(y' | y^*)} \pi_t(y^* | y^*) - \pi_t(y^* | y') \right]}$$

and the term  $\kappa(d_t) \frac{\pi_t(y_t|y)}{1+r}$  measures the loss in bond prices in case a default is decided.

- Case 2: The government starts the period in default. In case of no settlement

$$\begin{aligned} q_t(b, y) &= L(b_{t-1}, y_t, e_t, q_t, d_t(=1), d_{t-1}(=1)) \\ &= q_{t-1}(b, y) \end{aligned}$$

and if a settlement is achieved

$$\begin{aligned} q_t(b, y) &= L(b_{t-1}, y_t, e_t, q_t, d_t(=0), d_{t-1}(=1)) \\ &= q_{t-1}(b, y) + \frac{\pi_t(y^* | y) \kappa(1-h)}{1+r} \end{aligned}$$

the last term at the right hand side measures the gain in the bond price menu of the settlement, note that it is decreasing with the haircut.

The main qualitative results of this paper can also be shown with alternative specifications. The appendix replicates the graphs of the paper with the perceived law of motion  $L(b_{t-1}, y^*, e_t, q_t, d_t, 1) = q_t$ . This case is an opposite extreme of the rational expectations case. With rational expectations the government is completely aware of the effect of its decisions on the bond price menu and in the other case it erroneously believes its actions will not affect in any way the menu. This difference implies that the optimal policies vary among cases because the government with rational expectations will include "reputational" considerations when deciding the level of debt, repayment and the settlement. This mechanism is not the main driver of the points highlighted in this paper so I will not study it in depth. See Rojas 2013 for an analysis of this mechanism and its implications for the default episodes outcomes and its empirical support.

## Chapter 2

# Expectations Formation and Investment During Recessions

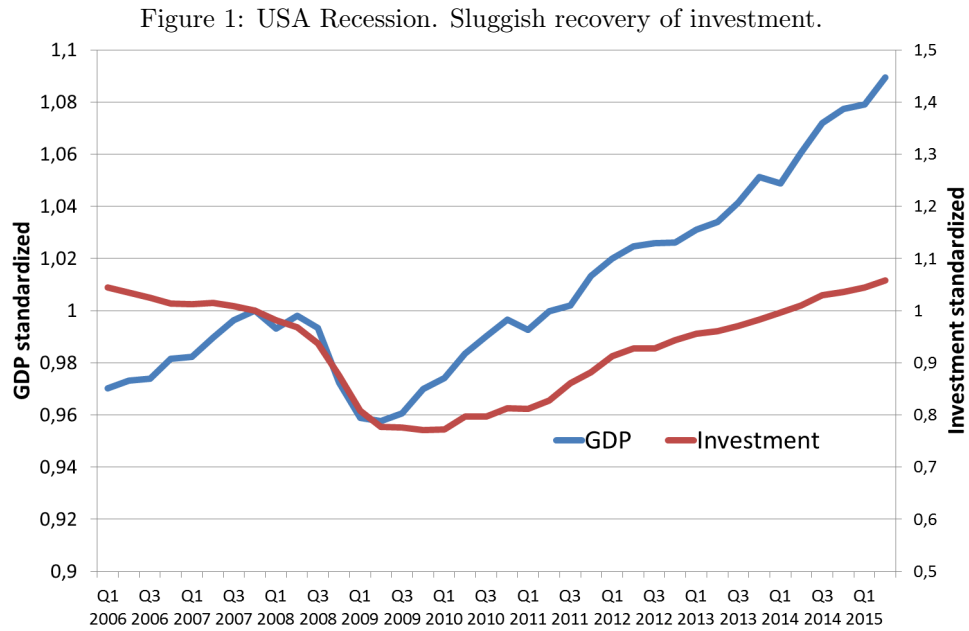
### **Abstract**

A standard assumption in macroeconomic modeling is that economic agents may have limited information but perfect market knowledge. I relax this assumption and study an environment where firms do not have perfect market knowledge and therefore form expectations with a simplified model of the economy. I study the investment decision of firms at the extensive margin along the business cycle, I show that during recessions firms tend to delay investment due to "pessimism": their subjective beliefs underestimate the potential gains of the investment project. The mechanism proposed helps to explain a stylized fact highlighted in this paper: After recessions, investment is slower in recovering than GDP. Since the pessimism induces an inefficient delay of investment I study the potential gains of an investment subsidy and a counter-cyclical corporate tax policy. If the government cannot distinguish between a lack of investment driven by pessimism and one by high technological risk, a counter-cyclical corporate tax policy can screen pessimistic firms and provide incentives to invest. On the other hand, an investment subsidy would push firms to invest in both cases, leading to excessive risk taking.

# 1 Introduction

From a Schumpeterian perspective, an economic crisis serves the purpose of accelerating the creative destruction process. Lower demand and lack of resources generate harsh competition for market shares. Many firms fail and jobs are destroyed. Once the economy starts to recover, the creative process should peak: as demand increases, new firms enter and the surviving incumbents grow. Nevertheless, I document that investment tends to recover sluggishly after a recession and takes more time to reach pre-recession levels than consumption does.

The great recession in the US is a clear example of the pattern of investment and GDP that I document. Figure 1 plots the levels of quarterly real investment and GDP from 2006 to 2015 standardized at 1 for Q4 of 2007. It takes about 2-3 years for GDP to regain the pre-recession peak levels. On the other hand, investment is already falling before GDP decreased and we observe a sluggish recovery in investment. It stays low even when GDP is already recovering and it takes more time to regain previous levels.



In this paper I argue that during recessions investors can become overly pessimistic about the profitability of projects and postpone their investment decisions. We define pessimism as a situation in which the subjective beliefs of the investor assign to a project an expected profitability which is lower than the one given by the objective distribution of outcomes.

I assume that firms have imperfect knowledge of the structure of the economy and use a simplified model to form their expectations about future dynamics. Firms behave rationally, as the predictions

of the model they use are actually confirmed by what is observed along the equilibrium path; in other words, firms do not commit systematic mistakes that contradict their reference model<sup>1</sup>. The equilibrium notion behind this reasoning is the Self-Confirming equilibrium, as proposed by Fudenberg and Levine (1993) and Sargent (2001). Importantly for this setup, beliefs are disciplined solely by what is observed along the equilibrium path<sup>2</sup>.

Since during recessions firms tend to invest less, and consequently rarely observe the outcome of the available projects, we can expect firms' beliefs about profitability of projects during recessions to be potentially biased. The possible bias is necessarily pessimistic, since an optimistic bias would induce investment and thus reveal the misbelief. In conclusion, this setup allows for firms to consistently hold excessively pessimistic beliefs about the off-equilibrium payoffs that investing during recessions would yield.

In a rational expectations equilibrium pessimism cannot arise by definition. Firms have subjective beliefs that correspond to the objective beliefs conditional on their information set. Even beliefs about the payoff of actions that in equilibrium are never taken have to satisfy this "consistency" requirement. This imposes a rich knowledge of the firm about the structure of the economy, even beyond what could be through experience. Furthermore, the economic models that are evaluated at the rational expectations equilibrium have the robust prescription that investment recovers faster than GDP. This goes from RBC models as in King and Rebelo (2000) to models with financial frictions and credit shocks such as in Khan and Thomas (2013). The sluggish pattern of investment calls for additional mechanisms that explain why investors do not react with higher investment as strong as what would be prescribed by the technological growth and the financial constraints.

The analysis in the paper has a simple event at its core: during a recession a firm decides to delay the investment to develop a new good because it expects low demand. These expectations are pessimistic. If it had invested it would have faced a higher demand. Once time passes and the firm eventually decides to invest, it faces a demand that is consistent with its expectations. Therefore, the pessimism during the recession is not revealed to the firm. I present a general equilibrium model and evaluate under which circumstances this event occurs.

The model consists of an economy where firms competing monopolistically can invest to produce a higher quality version of their differentiated product. The demand side is formed by consumers that have a valuation of quality that is increasing in the level of consumption. In this setup, changes in the income distribution affect the aggregate demand for different quality levels and therefore the profitability of investments in the development of a new good. The model is an extension of Fajgelbaum, Grossman, and Helpman (2011) to a dynamic model with capital and subjective beliefs.

Firms hold a model of the economy that consists in a simple forecasting scheme: a linear mapping from their available information to the profitability of launching a new quality level. Firms' model

<sup>1</sup>For example, their model of the economy predicts a correlation coefficient between current GDP and future demand of 0.3, and this is what they will effectively observe along the equilibrium path.

<sup>2</sup>One could describe this type of belief formation as a deductive process in which firms face a class of possible models of the economy and select the one that best explains observed outcomes. The selected model is then used to predict future outcomes for every possible situation the firm may face.

satisfies a consistency requirement: the moments implied by the model are satisfied in equilibrium. I draw on the work by Anderson and Sonnenschein (1985) on the model-consistent equilibrium concept and Sargent (2001) on Self-Confirming Equilibria.

In this novel theoretical setup, I illustrate how recessions that are accompanied by changes in the income distribution can generate inefficient delays in investment due to pessimistic beliefs. This feature is stronger for goods at the highest and lowest quality levels; meaning goods mostly consumed by the tails of the income distribution.

The intuition for the emergence of pessimism is as follows: the demand for a given quality level is the aggregation over consumers that are heterogeneous in their wealth and tastes. Therefore, At a given quality level, the observed demand does not reveal how many of these consumers are willing to pay more for an increase in quality. On average, along the equilibrium path, the firm knows how changes in the relative demands of various quality levels map in the potential demand of a new good. But when there are also changes in the income distribution this mapping might be biased and generate pessimistic beliefs.

On the policy side, I propose that if the government cannot distinguish between pessimism and high technological risk a counter-cyclical corporate tax can screen pessimistic firms and provide incentives to invest while an investment subsidy would push firms to invest in both cases implying excessive risk taking.

The paper is organized as follows: In section *II*, I document the sluggish recovery of investment after recessions and in section *III* I discuss the related literature. Section *IV* presents a simplified 2-period version of the model that illustrates how pessimism can emerge and the equilibrium concept. Section *IV* extends the simple model to a  $T$  period case to highlight the timing and potential delay of investment generated by the mechanism. In section *V* the full model with capital accumulation and innovation at multiple quality levels is presented and the technological and financial recessions characterized. Section *VI* contains the policy analysis and finally section *VII* concludes.

## 2 Investment Recovery After a Recession

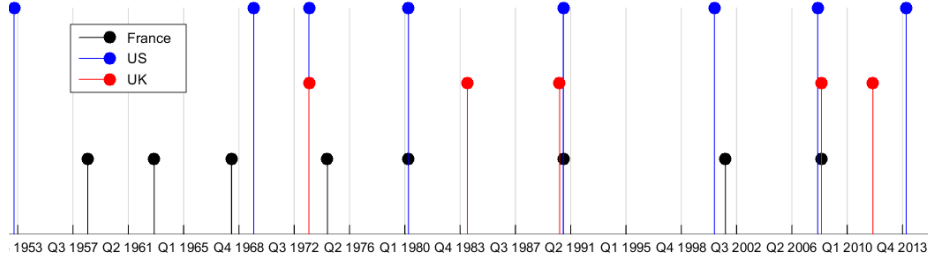
The pattern shown in Figure 1 is not a distinctive feature of the great recession in the US. In this section I provide evidence that investment does tend to recover slower than consumption and even slower than GDP using data from the US, UK and France.

I define the starting date of a recession the first quarter in which real GDP growth is negative after 2 years with positive quarterly growth rates. Following this criteria I identify 5 recessions in the UK and 8 recessions in the US and France. Figure 2 shows the time line with the identified recessions.

3 plots the number of quarters it took for GDP and investment to reach at least the levels each had one quarter before the starting date of the recession. The dark continuous line is the 45 degree line. Every point above this line represent a recession where investment recovery was slower than



Figure 2: Timing of the Starting Points of Recessions



Country	Quarter the recessions started
US	Q2 1953; Q3 1969; Q2 1973; Q1 1980; Q3 1990; Q4 2000; Q4 2007; Q4 2013
UK	Q2 1973; Q1 1984; Q2 1990; Q1 2008; Q3 2011
France	Q3 1958; Q1 1963; Q2 1968; Q4 1974; Q2 1980; Q4 1990; Q4 2001; Q2 2008

GDP. We see that most of the observations lay in this region. Investment does not tend to lead the recoveries after a crisis.<sup>3</sup>

The fact I document impose a challenge to our models of the economy. Most of the literature focuses on highlighting the higher volatility of investment relative to consumption or GDP. King and Rebelo (2000) for instance, show that for the US investment is three times as volatile as GDP. The notion of high volatility should not be confused with fast adjustments before and after the recessions. High volatility can also come from prolonged and persistent deviations from the mean. In fact King and Rebelo (2000) also find that the HP-detrended investment is as persistent as GDP.<sup>4</sup>

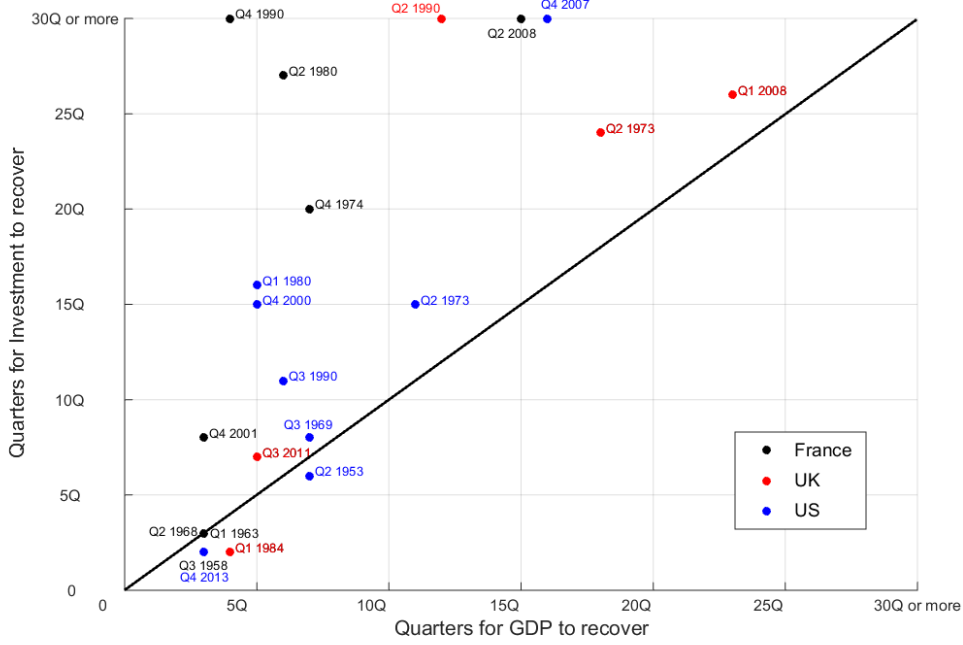
### 3 Related Literature

Within the rational expectations equilibrium approach Van Nieuwerburgh and Veldkamp (2006), and more recently Fajgelbaum, Schaal, and Taschereau-Dumouchel (2014), propose that leaning

<sup>3</sup>Note that for the case of GDP the reference level is, by construction, the highest GDP level over the last two years before the crisis. For investment this is not the case. Typically, investment is already at lower levels than the maximum reached over the years previous to the recession. We want to emphasize that this election should push the results in favor of a faster computed recovery of investment relative to output; that has the reference level at the previous maximum.

<sup>4</sup>King and Rebelo (2000) present an RBC model that is able to reproduce US data with small technological shocks introducing capital utilization in the model. With capital utilization they do not have to resort to large technological shocks, that potentially generate technological regress. The prescription of their model is standard in the business cycles literature, Investment and GDP recover contemporaneously. The stylized fact we document is partly "cleaned out" of their data as they use Hodrick-Prescott detrended series for all variables. The sluggish recovery of investment we document is partly absorbed by the trend obtained using the HP filter.

Figure 3: Length of the Recovery Spell for GDP and Investment



asymmetries during the business cycle partly explain the asymmetry observed in the behavior of GDP. In their model agents have to learn about productivity and during recessions there is less information available to update beliefs due to lower economic activity. This feature generate a slower but still, contemporary, recovery of GDP and investment. In my analysis a similar intuition goes through, there is less information about the assessment of investment projects during recessions. The main departing points of my approach with respect to Van Nieuwerburgh and Veldkamp (2006) are that expectations are formed about potential demand, and investment is also decided at the extensive margin. Some firms do not invest and consequently do not observe demand, this differences allow for the existence of a self-confirming equilibrium where there can be inefficient delays in investment driven by expectations.

Bloom (2009) on the other hand studies delays in investment motivated by "uncertainty" shocks. As the volatility of fundamentals increase, the firms expand their inaction region and decide not to invest in a wider set of situations. At the edge of the recession the uncertainty can be high and investment delayed, but once the economy starts to recover investment spikes and lead the recovery.

Departing from Bloom (2009), the mechanism I propose generates further delays in investment even when uncertainty has decreased and can sustain a slow recovery of investment during recoveries with low measured volatility.

A contrasting analysis to mine that relates pessimism by firms and low investment is Kiyotaki (1988). Although he does not focus on fluctuations, it is shown that there could be multiple equilibria where firms expectations of demand affect their investment decision and eventually demand itself. Pessimist firms decide to have a low level of investment. Opposite to the concept of pessimism considered here, firms are right about expecting low demand as in equilibrium they face effectively low demand. The equilibrium is self-fulfilled and the outcome of the investment project is always observed. With the notion of pessimism I propose a firm in a self-fulfilled equilibrium is not pessimistic, as its beliefs coincide with the objective expectations.

The difference between a self-fulfilled recession and a self-confirmed seems subtle but can have important policy implications. The self-fulfilled recession is a problem of coordination while the self-confirmed case can be framed as a problem of lack of knowledge or experimentation. The policy implications to solve each recession are consequently different. Kiyotaki (1988) proposes an investment subsidy to coordinate firms in the high investment equilibria. On the other hand, I argue that in the context of a self-confirming equilibrium where only some firms are not investing out of pessimistic beliefs and the rest because of technological risk such policy can lead to excessive risk taking. A policy that screens the potentially pessimistic firms can be superior by inducing experimentation rather than coordination.

I build on the idea of a self-confirming equilibrium in a competitive economy or as Gaballo and Marimon (2015) define it a Strong self-confirming equilibrium and focus on subjective beliefs about market outcomes by firms. Gaballo and Marimon (2015) propose how a self confirmed equilibrium can sustain a crisis. In their setup the banks have subjective beliefs about the investment possibilities of firms and consequently the type of projects firms will choose to implement if a loan is conceded. Pessimistic banks expect firms to take risky projects and therefore raise the interest rate. The only profitable project for firms with the high interest rate is indeed a risky one. The unobserved counterfactual is that if a single bank offers a low interest rate the firm will optimally select a safe investment, something the bank subjectively beliefs will not happen.

Gaballo and Marimon (2015) beliefs are formed about the technological frontier of projects. In this paper I focus on beliefs about market outcomes rather than the technological possibilities. We think this type of uncertainty is important during a crisis like the great recession where the behavior of demand might entail a greater increase in uncertainty than the technological process itself.

On the policy side Gaballo and Marimon (2015) show how a credit easing policy can be an optimal policy to exit the crisis. Just as mentioned for the policy suggested by Kiyotaki (1988) this could lead to excessive risk taking in the economy I analyze. The reason is that both, investment subsidies and credit easing are aimed at reducing the cost of investment. This will affect both firms that would not invest because of technological risk or low expectations about demand. The solution I propose is to provide the incentives to invest from the income side by changing the corporate income

tax. Change in the corporate income tax affects different firms depending on their expected profits and the likelihood to succeed innovating. This way a lower tax for low profits will push pessimistic firms to invest while it might not induce firms with high technological risk to invest.

## 4 The Basic Model

In this section I use a simple 2-period model to illustrate and explain the key elements of the analysis: The innovating firms and their expectations formation; the consumers with a taste for quality; and the equilibrium concept. Then I extend the basic setup to a T-period economy to show the timing and potential delay of investment.

### 4.1 Setup

The model is a dynamic version of Fajgelbaum, Grossman, and Helpman (2011) with subjective beliefs. The economy lasts for two periods and is populated by firms and consumers. There are two types of goods, a basic good and a differentiated good. The basic good is an homogenous good produced by a competitive firm with constant returns to scale. The differentiated good has many versions, as each firm in this sector owns a differentiated variety that can be produced at multiple quality levels. The firms start period one with two quality levels variety; compete monopolistically; and decide whether to develop a higher quality version of its variety to be sold in the second period.

Households consume only one unit of the differentiated good and their decision is to select which version to buy. There are two sources of heterogeneity in the population: the preferences over the differentiated good and labor productivity. Both are random and are drawn at the beginning of the first period. The distribution of skills in the population is also random and drawn in period one. Then there are idiosyncratic and aggregate shocks.

#### The Basic Good Producer

There is a representative firm that produces the basic good using effective labor as the only input. Let  $y^i$  denote worker  $i$  productivity, then the production function of the firm is

$$Y = \int_{\Psi} y^i dh$$

where  $\Psi$  is the set of workers hired by the firm where each worker supplies one unit of labor. The firm has constant returns to scale and is competitive in the goods market and in the labor market. Taking the basic good as the numeraire in our economy we have that in equilibrium the wage is equal to 1 and consequently the income of worker  $i$  is equal to  $y^i$ .

### Taste for Quality Preferences

The specification of preferences is based on Fajgelbaum, Grossman, and Helpman (2011). The utility level of an individual  $i$  that consumes a level  $z$  of the basic good and the differentiated variety  $j$  with quality level  $q$  is

$$u^i(z, j, q) = zq + \varepsilon_{j,q}^i$$

where  $\varepsilon_{j,q}^i$  is the idiosyncratic valuation of variety  $j$ , at quality level  $q$ , by household  $i$ . Let  $\varepsilon^i$  be the vector containing all the idiosyncratic taste components of household  $i$ . The preferences impose complementarity between the quality of the differentiated good and the quantity consumed of the basic good. This implies that there is a higher valuation of quality by those that consume higher amounts of the homogenous good  $z$ .

Following Fajgelbaum, Grossman, and Helpman (2011) we take the probability distribution of  $\varepsilon^h$  to be given by a generalized extreme value (GEV) distribution. The distribution is given by:

$$G_\varepsilon(\varepsilon) = \exp \left[ - \sum_{q \in Q} \left( \sum_{j \in J_q} e^{-\varepsilon_{j,q}/\theta_q} \right)^{\theta_q} \right] \quad (1)$$

where  $Q$  is the set of all quality levels available,  $J_q$  is the set of varieties available at quality level  $q$  and  $\theta_q \in [0, 1]$  for all  $q \in Q$  is the dissimilarity parameter. The higher  $\theta_q$  the less correlated are the shocks at quality level  $q$ , and consequently a higher monopoly power at quality level  $q$ . Note that the shocks for the same variety  $j$  at it's different quality levels are not correlated.<sup>5</sup> I also suppose that the shocks are i.i.d across households.

### Household Decision and Heterogeneity

There is a continuum of households  $i \in [0, 1]$  of mass one. Household  $i$ 's income  $y^i$ , is used to buy the basic good and one unit of the differentiated good. The budget constraint is then

$$y^i = z + \sum_{q \in Q} \left( \sum_{j \in J_q} I_{j,q}^i p_{j,q} \right)$$

where  $p_{j,q}$  is the price of variety  $j$  at quality level  $q$ ; and  $I_{j,q}^i$  is the indicator function that takes the value of one only if the variety consumed by household  $i$  is  $j$  at quality level  $q$ . Households cannot save and therefore the two period household problem is in essence two consecutive static

---

<sup>5</sup>This property greatly simplifies the optimal pricing strategy of firms. We discuss the role of this assumption in the section (4.1)

problems.<sup>6</sup> The household problem at each  $t = 1, 2$  is to solve

$$\begin{aligned} \max_{z, \{I_{j,q}\}_{j \in J, q \in Q}} \quad & u^h(z, j, q) \\ \text{s.t.} \quad & y^h = z + \sum_{q \in Q} \left( \sum_{j \in J_q} I_{j,q}^h p_{j,q} \right) \end{aligned} \quad (2)$$

Let  $H(y)$  be the cumulative distribution of  $y^i$ .  $H(y)$  is random and drawn in period one from the distribution  $P(H(y))$ . The realization of the income distribution is the only “shock” that this economy faces.

### Aggregate Demand

We denote  $z^h(\{p_{q,t}^j\}_{i \in J, q \in Q_{j,t}}, y^h)$  to be the demand of household  $h$  for the basic good at time  $t$ . This demand is the solution of the household problem (2) and is consequently a function of the prices of the available differentiated good versions  $\mathbf{p}_t = \{p_{q,t}^j\}_{i \in J, q \in Q_{j,t}}$  and the household income  $y^h$ . Aggregate households demand for  $z$  is then

$$Z(\{p_{q,t}^j\}_{i \in J, q \in Q_{j,t}}, H(y)) = \int_0^1 z^h(\{p_{q,t}^j\}_{i \in J, q \in Q_{j,t}}, y^h) \partial H(y)$$

For the differentiated good, as Fajgelbaum, Grossman, and Helpman (2011) show, the aggregate demand system is a nested logit. Let  $\mathbf{p}$  be the vector whose entries are all the prices of the available versions of the differentiated good; and  $\rho_{j,q}^y(\mathbf{p}, \{J_q\}_{q \in Q})$  be the fraction of households with productivity  $y$  that buy variety  $j$  at quality level  $q$ . Then aggregate demand corresponds to the aggregation of  $\rho_{j,q}^y(\mathbf{p}, \{J_q\}_{q \in Q})$  over the income levels as follows

$$D_{j,q}(\mathbf{p}, H(y), \{J_q\}_{q \in Q}) = \int \rho_{j,q}^y(\mathbf{p}, \{J_q\}_{q \in Q}) \partial H(y) \quad (3)$$

each fraction is weighted by the density of population at each income level.<sup>7</sup>

The fraction  $\rho_y(j, q)$  is determined by the prices of all versions (variety-quality duple) of the differentiated good and the parameters that describe preferences as follows

$$\rho_{j,q}^y(\mathbf{p}, \{J_q\}_{q \in Q}) = \phi_{j|q}(\mathbf{p}, J_q) \lambda_q(y, \mathbf{p}, \{J_q\}_{q \in Q}) \quad (4)$$

where

$$\phi_{j|q}(\mathbf{p}, J_q) = \frac{e^{-p_j q / \theta_q}}{\sum_{l \in J_q} e^{-q p_l / \theta_q}} \quad (5)$$

<sup>6</sup>I impose this assumption to focus on the expectation formation of firms, note anyway that it is not highly restrictive given the fact that households do not face income uncertainty as the level of productivity for the two periods is the same.

<sup>7</sup>Recall that there is a unit mass of households and they demand only a one unit of the differentiated good and therefore we only need the fraction  $\rho_y(j, q)$  and the distribution of agents over  $y$  to compute aggregate demand.

is the fraction of consumers who buy variety  $j$  out of those who choose quality level  $q$ , and

$$\lambda_q(y, \mathbf{p}, \{J_q\}_{q \in Q}) = \frac{\left[ \sum_{j \in J_q} e^{(y-p_j)q/\theta_q} \right]}{\sum_{\omega \in Q} \left[ \sum_{j \in J_q} e^{(y-p_j)\omega/\theta_\omega} \right]^{\theta_\omega}} \quad (6)$$

is the fraction of agents with income  $y$  that choose quality level  $q$ . Note that  $\theta_q$ , the dissimilarity parameter, determines the price elasticity of aggregate demand. The higher  $\theta_q$  the lower is the effect on aggregate demand of a change in the price, i.e higher market power for the firm.

### Firms' Structure and Pricing

There are  $J$  firms in the economy owned by risk neutral investors. Denote  $Q_{j,t}$  the set of quality levels available for firm  $j$  at time  $t$ . Each firm is born with two quality levels  $Q_{j,1} = \{q_L, q_M\}$ , where  $q_L < q_M$ , and a corresponding unit production cost  $\{c_{q_L}, c_{q_M}\}$ . At the end of the first period the firm can develop a new quality level  $q_H$  that satisfies  $q_H > q_M$ , the cost of the innovation project is  $f_q$  units of the basic good. Once the quality level  $q$  is developed it can be produced at the unit cost  $c_{q_H}$ .

Firms seek to maximize their discounted stream of profits and are not credit constrained. I suppose that firms know the elasticity of demand as this elasticity is constant for all possible productivity/income distributions  $H(y)$ . Then it follows that the optimal pricing strategy of firms is in essence a static problem: taking as given the quality levels available  $Q_{j,t}$  at  $t$  the firm chooses prices to maximize profits at  $t$ .<sup>8</sup> The problem corresponds to

$$\max_{\{p_q^j\}_{q \in Q_{j,t}}} \sum_{q \in Q_{j,t}} E^j \left\{ D_{j,q} \left( \mathbf{p}_{t+1}, H(y), \{J_q\}_{q \in Q_{t+1}} \right) \right\} [p_q^j - c_q]$$

$E^j\{D_{j,q}(\cdot)\}$  denotes firm  $j$  expectations of demand for variety  $j$  at quality level  $q$ .

With a large number of firms  $J$  and using the fact that the idiosyncratic taste shocks are not correlated at the firm level, we have that the effect of a change in the price of the quality level  $q$  ( $p_q^j$ ) has a negligible effect on the demand for the good with quality  $q$ . Therefore, the previous problem can be split into many independent pricing problems, one for each quality level of the firm as follows

$$\sum_{q \in Q_{j,t}} \left( \max_{p_q^j} E^j \{ D_{j,q}(\cdot) \} [p_q^j - c_q] \right)$$

we can focus then on the pricing problem for a given variety  $j$  and quality level  $q$  to fully characterize the whole pricing problem. The problem is simply stated then as

$$\max_{p_q^j} E^j \{ D_t(j, q) \} [p_q^j - c_q] \quad (7)$$

---

<sup>8</sup>Firms have no incentives to use prices to extract information about the aggregate demand structure as the constants elasticity implies that changes in prices do not convey any additional information.

The optimal price is a markup over the marginal cost that depends on the elasticity of demand. The price that solves (7) is

$$p_q = c_q + \frac{\theta_q}{q} \quad (8)$$

where the markup ( $\theta_q/q$ ) is fixed, constant and only depends on the quality level but not on demand. We see how a greater value of the dissimilarity parameter  $\theta_q$  increases the mark-up. We also obtain then that for the markup to increase with quality it is necessary that  $\theta_q$  increases with quality.

### Innovation decision

The expected discounted profits the firm obtains if it develops the quality level  $q_H$  for its variety are given by

$$E^j \{ \pi_j^{q_H} \} = \frac{1}{R} \left[ E^j \left\{ D_{j,q_H} \left( \mathbf{p}_{t+1}, H(y), \{J_q\}_{q \in Q_{t+1}} \right) \right\} (p_{q_H} - c_{q_H}) - f_{q_H} \right] \quad (9)$$

where  $p_{q_H}$  is given by equation (8) and  $E^j[\cdot]$  represents the subjective beliefs of firm  $j$ . Then, if the expected net gain of the investment  $E^j \{ \pi_j^{q_H} \}$  is positive it is optimal for the firm  $j$  to invest in the new quality. The source of uncertainty for the investment decision is the future demand for the not yet developed quality.

The firms forms expectations about future demand based on their sales during the first period and the aggregate economic activity measured by the total production of the basic good in the economy. Formally we set subjective expectations as

$$E^j \left\{ D_{j,q_H} \left( \mathbf{p}_{t+1}, H(y), \{J_q\}_{q \in Q_{t+1}} \right) \mid X_j \right\} = M(X_j; \beta_j) \quad (10)$$

where  $M(X_j; \beta_j)$  is a function of vector  $X_j$  and parametrized by  $\beta_j$ . the vector  $X_j$  contains the total output  $Y$  and the first period sales of firm  $j$ . Note that with the general formulation of  $M(X_j; \beta_j)$  we have not discarded the possibility that the subjective expectations coincide with the rational expectations. In section 4.4 we take an specific measure  $P$  over the distribution for endowments  $H(y)$  for which, even in the case of full information,  $X_j$  are sufficient statistics to form the rational expectations.

Replacing 10 in 9 we have that the firm will invest to innovate and produce a new quality level if

$$M(X_j; \beta_j)(p_{q_H} - c_{q_H}) - f_{q_H} \geq 0$$

this is the optimal investment strategy for firm  $j$ .



### Market clearing

Let  $l^j$  be the firm  $j$  demand for inputs. It is equal to the unit cost of the versions they sell  $c_q$  times the amount produced of each each variety; therefore we have

$$l^j = \sum_{q \in Q_{j,t}} c_q D_t(j, q)$$

and aggregating for all firms  $j \in J$  we get the aggregate demand for inputs  $L$  that is equal to

$$L = \sum_{j \in J} l^j$$

The total production of the basic good in the economy is given by  $Y = \int_0^1 y^h H(y)$ . Then, the resource constraint is

$$Y = Z + L + \Pi \quad (11)$$

where  $\Pi$  is the aggregate level of profits and is given by

$$\Pi = \sum_{j \in J} \pi_j = \sum_{q \in Q_{j,t}} D_{j,q}(\cdot) \left[ \frac{\theta_q}{q} \right]$$

we see that profits are in the basic good units. They can be used to finance the investment on innovation or to pay dividends.

## 4.2 Equilibrium definitions

In this section I define the equilibrium with subjective beliefs and the rational expectations equilibrium. Then I present how subjective beliefs are disciplined and discuss how it corresponds to known equilibrium concepts. Finally I proposed an equilibrium concept that seeks to explicitly differentiate between potential biases along the equilibrium path and outside the equilibrium path.

### Equilibrium with subjective beliefs

For each possible realization of the endowment distribution  $H(y)$  with positive probability according to  $P$  an equilibrium is a vector of available qualities  $\{Q_{j,t}\}_{j \in J, t=1,2}$ ; a vector of prices of the differentiated good  $\{p_{q,t}^j\}_{i \in J, q \in Q_{j,t}, t=1,2}$ ; an allocation of the differentiated good  $\{I_{j,q,t}^h\}_{j \in J, q \in Q_{j,t}, t=1,2, h \in (0,1)}$  and consumption of the basic good  $\{z^h\}_{h \in (0,1)}$ ; a vector of aggregate demands  $\{D_{j,q}\}_{j \in J, q \in Q}$ ; such that given a realization of the taste shocks  $\varepsilon$ , and the functions  $\{M(X_j; \beta_j)\}_{j \in J}$  satisfy:

1.  $Q_{j,1} = \{q_L, q_M\}$  for all  $j \in J$  and  $Q_{j,2} = \{q_L, q_M, q_h\}$  if and only if  $M(X_j; \beta_j) \geq \frac{f_{q_H}}{p_{q_H} - c_{q_H}}$ , otherwise  $Q_{j,2} = \{q_L, q_M\}$ .
2. The prices  $\{p_{q,t}^j\}_{i \in J, q \in Q_{j,t}, t=1,2}$  are set accordingly to (8).

3. The indicators  $\{I_{j,q,t}^h\}_{j \in J, q \in Q_{j,t}, t=1,2}$ , and consumption  $z^h$  solve the problem (2) for each  $h \in [0, 1]$ .
4. The demands  $\{D_{j,q}\}_{j \in J, q \in Q}$  are given by 3, 4, 5 and 6.
5. The resource constraint (11) holds.

### Rational expectations equilibrium (REE)

The REE has to satisfy the additional condition

6. The subjective expectations  $M(X_j; \beta_j)$  coincide with the conditional expected value

$$E_P \{D_{j,q} \mid X_j\}$$

The condition requires that the subjective expectations coincide with the objective expected value. Note that this has to hold even for firms that do not decide to invest and therefore outside the equilibrium path. We will refer from now on to  $E_P \{D_{j,q} \mid X_j\}$  as the objective beliefs.

### 4.3 Subjective Beliefs

The definition of the equilibrium with subjective beliefs poses no constraint on  $M(X_j; \beta_j)$  and up to this point it could be set completely arbitrary. I now proceed to give structure to such beliefs to discipline how the agents form expectations. The expectations  $M(X_j; \beta_j)$  could in principle nest the rational expectations beliefs. Given the functional form of  $M(X_j; \beta_j)$  each value of the vector  $\beta$  determine a different forecast function. We define two types of beliefs according to two criteria to select  $\beta$ .

#### Parametrized Expectations

Given the distribution  $P(H(y))$  over the possible distributions  $H(y)$  these expectations are constructed with  $\hat{\beta}$  such that it solves

$$\hat{\beta} = \arg \min_{\beta} E_{P(H(y))} L(M(X_j; \beta_j), E_P \{D_{j,q} \mid X_j\})$$

where  $L$  is a loss function that measures the distance between the subjective and objective beliefs. Note that any differences, as defined by  $L$ , between  $M(X_j; \beta_j)$  and  $E_P \{D_{j,q} \mid X_j\}$  are minimized according to the measure  $P$ . If  $M(X_j; \beta_j)$  nests  $E_P \{D_{j,q} \mid X_j\}$  then, if  $L$  is the euclidean distance, for all distributions  $H(y)$  with positive density they will coincide. If  $M(X_j; \beta_j)$  does not nest  $E_P \{D_{j,q} \mid X_j\}$  then the differences between the subjective and objective beliefs are the result of misspecification.<sup>9</sup>

---

<sup>9</sup>den Haan and Marcet (1990) early contribution show how parametrized expectations can approximate the rational expectations and study the approximation errors.

### Self-Confirming Expectations

This expectations minimize the difference between objective and subjective beliefs but only using those events that belong to the equilibrium path and consequently are potentially observable. Since firms are forming expectations about demand, and is only observed if they decide to invest, then we fit the subjective expectations to the situations were firms do invest.

Formally,  $\hat{\beta}$  is selected accordingly with

$$\hat{\beta} = \arg \min E_{P(H(y)|H(y) \in S(M(\nu, Y; \hat{\beta})))} L(M(X_j; \beta_j), E\{D_{j,q} | X_j\})$$

where  $S(M(\nu, Y; \beta)) = \{H(y) \in P(H(y)); M(X_j; \beta) \geq M^*\}$  and  $M^*$  is the threshold to invest in the project. Therefore, the loss function is only evaluated at those events that are potentially observable when the firm invests and not otherwise. Note the self-referential definition of  $\hat{\beta}$ . It is such that it minimizes the loss function over the probability space characterized by itself. This comes from the fact that the selection of  $\hat{\beta}$  affects itself which outcomes are potentially observable. We could write it as the fixed point of the mapping

$$\mathcal{F}(\beta^*) = \arg \min_{\beta} E_{P(H(y)|H(y) \in S(M(X_j; \beta^*)))} L(M(X_j; \beta_j), E\{D_{j,q} | X_j\})$$

such that  $\hat{\beta}$  is the vector that satisfies

$$\hat{\beta} = \mathcal{F}(\hat{\beta}) \tag{12}$$

Agents in this case do inference about the remaining events using the functional form  $M$  and the parameters  $\hat{\beta}$ . This expectations could be obtained as the limit of learning from experience by agents that can only used their observed outcomes.

Here on top of the misspecification bias, if there is one, the agents expectations are affected by the “sample” selection bias. Even though misspecification might be negligible we emphasize that once it interacts with the fact that agents learn from a selected sample its implications can become relevant. Furthermore, even in the absence of misspecification, for a rich enough statistical model there could be multiple solutions to 12 and therefore potential disagreements in beliefs for non observed events.

### Self-confirming equilibrium (SCE)

The subjective expectations equilibrium with the Self-Confirming expectations is what Sargent (2001) defines as a Self-Confirming equilibrium. If we take the loss function  $L$  to be a quadratic loss function (euclidean distance), then the firms beliefs satisfy the following orthogonality condition

$$E_{P(H(y)|H(y) \in S(M(\nu, Y; \hat{\beta}^*))} \left\{ \left[ D_{j,q} - M(X_j; \hat{\beta}_j) \right] M(X_j; \hat{\beta}_j) \right\} = 0$$

If we set the beliefs to be a linear mapping of  $X_j$  of the form  $M(X_j; \beta_j) = \beta_j X_j$ , the orthogonality

conditions corresponds to

$$E_{P(H(y)|H(y) \in S(M(\nu, Y; \beta^*))} \left\{ \left[ D_{j,q} - \hat{\beta}_j X_j \right] X_j' \right\} = 0$$

in this case the firm identifies the correlation between the variables in their information set and future demand.

In a Self-Confirming equilibrium, even along the equilibrium path we could have that

$$M(X_j; \hat{\beta}_j) \neq E \{ D_{j,q} \mid X_j \}$$

as long as the orthogonality conditions are satisfied. Contrastingly in Fudenberg and Levine (1993) definition of a SCE this two objects have to be equal along the equilibrium path and can only be potentially different outside the equilibrium path. To relate the two definitions in my analysis I define a measure to evaluate how close are the subjective expectations to the objective expectations and state a weaker version of the Fudenberg and Levine (1993) definition.

Let the function  $R_\Omega(d)$  be defined as follows

$$R_\Omega(d) = Pr \{ L(M(X_j; \beta_j), E \{ D_2 \mid X_j \}) < d \mid H(y) \in \Omega \}$$

where  $Pr(A)$  stand for the probability of event  $A$ . The function  $R_\Omega(d)$  measures the set for which the distance between the objective and the subjective expectations is smaller than  $d$ , conditional on the fact that the productivity distribution lies within the set  $\Omega$ . Note that  $R_\Omega(0) = 1$  means that within the set  $\Omega$  the subjective beliefs coincide with the rational expectations.

In our case a SCE according to the Fudenberg and Levine (1993) definition happens if  $R_\Omega(0) \leq 1$  and  $R_\Upsilon(0) = 1$  for  $\Upsilon = \{ (H(y) \mid H(y) \in S(M(X_j; \beta_j^*))) \}$ . Therefore we have that subjective beliefs coincide with the objective beliefs along the equilibrium path but might not coincide in a positive measure set outside the equilibrium path.

### $\epsilon$ -Self-Confirming Equilibrium

Fudenberg and Levine (1993) refer to the notion of an  $\epsilon$ -consistent SCE. In this setup beliefs will be within some  $\epsilon$  distance of the equilibrium outcomes.<sup>10</sup> In my setup it is also possible to characterize if the beliefs are within an  $\epsilon$  distance from the equilibrium outcomes. Define an  $\epsilon$ -Self-Confirming equilibrium an equilibrium with subjective beliefs  $M(X_j; \beta_j^*)$  such that there exists an  $\epsilon > 0$  that satisfies  $R_\Upsilon(\epsilon) = 1$ , for  $\Upsilon = \{ (H(y) \mid H(y) \in S(M(X_j; \beta_j^*))) \}$ . Throughout the paper I am going to focus on the Sargent (2001) definition and check ex-post the minimum  $\epsilon$  such that the equilibrium found is also  $\epsilon$ -Self-Confirming.<sup>11</sup>

<sup>10</sup>This concept is obtained by having the beliefs of a player correspond to a game where there is a fraction  $\epsilon$  of "crazy" types. Where crazy types deviate from the equilibrium strategies.

<sup>11</sup>At this point, the reader familiar with Krusell, Smith, and Jr. (1998) might have spotted the parallelism with the approach followed here. In Krusell, Smith, and Jr. (1998) households have to make expectations about future

To emphasize the role of beliefs out of the equilibrium path and have a measure of the potential deviations of beliefs from the objective expectations I propose the notion of an  $\epsilon, \delta$ -Self-Confirming equilibrium.

#### $\epsilon, \delta$ -Self-Confirming Equilibrium

An  $\epsilon, \delta$ -Self-Confirming Equilibrium is a subjective beliefs equilibrium for which  $R_Y(\epsilon) = 1$  and  $R_\Omega(\delta) < 1$  where  $\epsilon \leq \delta$ .

Just as in an  $\epsilon$ -Self-Confirming equilibrium deviations along the equilibrium path ( $\mathcal{Y}$ ) are at most  $\epsilon$ . The novelty is that in this definition is required that the distance between the subjective and objective beliefs is not bounded above by  $\epsilon$  outside the equilibrium path, and in fact out of the equilibrium path the deviation is greater than  $\delta$ . The self-confirming equilibria I discuss in the next sections satisfy the requirements for a  $\epsilon, \delta$ -Self-Confirming Equilibria.<sup>12</sup>

#### 4.4 Equilibrium Objective and Subjective Beliefs with a Simple Class of Income Distributions $P(H(y))$ .

To illustrate how a SCE that generates a lack of investment can emerge and be different from the rational expectations equilibrium, I focus on a specific measure over the possible endowment distributions  $H(y)$  and define the subjective beliefs  $M(X_j; \beta_j)$  as a linear mapping.

Let  $H(y; Y, \sigma)$  be the distribution of a discrete random variable  $y$  with parameters  $Y$  and  $\sigma$  that determine the average productivity and the standard deviation. Specifically, we suppose that half of the population gets an endowment equal to  $y_L$  and the other half equal to  $y_H$  that are given by

$$\begin{aligned} y_L &= Y - \sigma \\ y_H &= Y + \sigma \end{aligned}$$

Now we can proceed to characterize  $P(H(y))$  with a joint probability distribution for  $\sigma$  and  $Y$ . This assumption greatly simplifies the analysis because aggregation is straightforward while we still keep the two dimensions we want to explore: Total income  $Y$  and inequality  $\sigma$ .

---

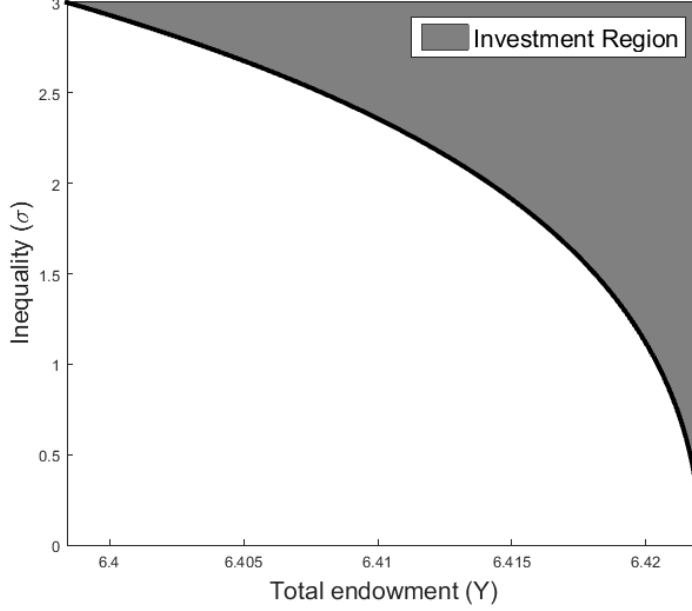
prices that is itself a function of the distribution of capital along the population. Just as the firms in my model have to form beliefs about demand that depend also on the whole income distribution. Krusell, Smith, and Jr. (1998) propose a simple forecasting function for households and compute an  $\epsilon$ -Self-confirming equilibrium where  $\epsilon$  is close to zero as shown by an  $R^2$  of the regressions of their model in beliefs almost equal to 1.

In our case the  $R^2$  of the beliefs to forecast demand is also close to 1 for the simple 2-period and above 0.9 in the full version of the model. Nevertheless, this paper illustrates that characterizing beliefs about equilibrium outcomes leaves out a key element that can determine agents decisions: "wrong" beliefs out of the equilibrium path. In particular, how those beliefs tend to be pessimistic and determine the decisions of households along the equilibrium path.

Krusell, Smith, and Jr. (1998) refer to the  $\epsilon$ -Self-Confirming equilibrium in their model as an approximate equilibrium. This reflects the notion that the equilibrium approximates the rational expectations equilibrium. I illustrate that this does not hold in general and there can be substantial differences between both.

<sup>12</sup>A stronger version of the definition would be that there exists a  $\delta$  such that  $R_Y(\epsilon) = 1$  and  $E_{\Omega/Y} \{L(M(\nu, Y; \beta), E\{D_2 | \nu, Y\})\} > \delta$ . In this definition the deviation of expectations is bounded below by  $\delta$  outside the equilibrium path.

Figure 4: Optimal investment rule with objective expectations



With this parametrization we proceed to the characterization of the objective and the subjective beliefs.

#### 4.5 Objective beliefs

Figure (4) shows the regions on the  $\sigma, Y$  plane where the firms decide to invest in the REE. The investment region is at the top right, meaning that high aggregate endowment  $Y$  and inequality  $\sigma$  make more profitable investments to produce the high quality goods. This follows from the fact that the innovation is done at the highest quality level and therefore the potential consumers are mostly the top of the income distribution. So higher mass at high levels of income makes more profitable the production of a high quality good.

#### 4.6 Subjective Beliefs

Subjective beliefs of firm  $j$  are a linear function of observed demand  $(D_{j,q_L}, D_{j,q_M})$  and the aggregate production of the basic good  $Y$  as follows

$$M^j(D_{j,q_L}, D_{j,q_M}, Y; \beta) = \beta_0 + \beta_1 D_{j,q_L} + \beta_2 D_{j,q_M} + \beta_3 Y \quad (13)$$

where  $\beta = (\beta_0, \beta_1, \beta_2, \beta_3)$  are parameters. The subjective expectations are a function of the same two variables as the rational expectations, the only difference comes from the functional form.

**Expectations based only on  $Y$** 

To illustrate the nature of the misspecification we start with case where we set  $\beta_1 = \beta_2 = 0$  and select  $\beta_0, \beta_3$  following the Self-Confirming expectations criteria. In this case the forecast function of demand is of the form

$$M^j(Y; \beta) = \beta_0 + \beta_3 Y$$

and therefore the optimal investment strategy is characterized by a threshold  $\bar{Y}$  such that

$$\pi(j, q_H) = 0$$

the optimal investment rule depends on the sign of  $\hat{\beta}_3$ . If  $\hat{\beta}_3 > 0$  then expected demand increases with  $Y$ , and it is optimal to invest if  $Y > \bar{Y}$ . Since we have that demand  $(D_{j,q_L}, D_{j,q_M})$  itself depends on  $Y$ , it is not obvious from the RE solution which value corresponds to  $\beta_3$  in the Self-Confirming expectations. The correlation of  $Y$  with demand depends on the joint distribution of  $(Y, \sigma)$ . To illustrate the type of investment strategy that emerges with the subjective beliefs, let  $P(Y, \sigma)$  be such that  $Y$  and  $\sigma$  are independent and have a uniform distribution for  $Y$  and  $\sigma$ .

The subject expectation (SE) threshold  $\bar{Y}$  is found and the corresponding parameter values. For the parametrization of  $H$  we have that  $\beta_2 > 0$ , the optimal investment strategy is to invest if  $Y > \hat{Y}$ . In figure 5 we plot the threshold for the subjective expectations SE and for the REE. We see that there are two regions where the investment strategies do not coincide  $A$  and  $B$ . In region  $A$  it is optimal to invest but the firms with the subjective expectations do not do it. On the other hand, in region  $B$  with the REE it is not optimal to invest but firms with subjective expectations do decide to invest and therefore face negative profits.

**Self-Confirming Equilibrium  $M(X_j; \hat{\beta}_j)$** 

Now we go back to the original specification (13) and compute the value  $\beta$  that correspond to the Self-Confirming expectations. Compared to the previous case the disagreement regions are smaller and in both cases we have the type  $A$ . The SE threshold is above the RE threshold and therefore in the gray region firms with subjective beliefs do not invest while along the RE equilibrium they would do it. The disagreement is minor and happens both at high GDP and low inequality and at low GDP and high inequality. Note that the two lines are almost identical for the centered values of  $Y$  and  $\sigma$ , those values are just the most likely according to  $H(Y, \sigma)$  and just reflects the nature of how the  $\beta$  is selected.

In Figure 7 the deviation between the objective beliefs and the subjective beliefs is presented for two cases. Case 1 refers to the parameterized expectations and simply captures the difference induced by the misspecification of beliefs. Case 2 refers to the SCE case. The positive levels represent pessimistic beliefs for a given level of  $Y$ . Misspecification induces pessimism at high and low levels of  $Y$  (case 1) but in the SCE (case 2) the fact that beliefs satisfy the orthogonality conditions only

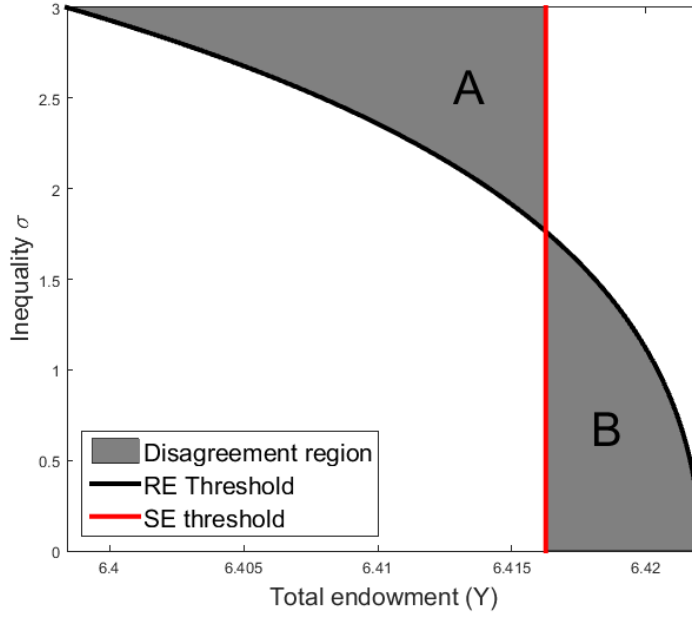
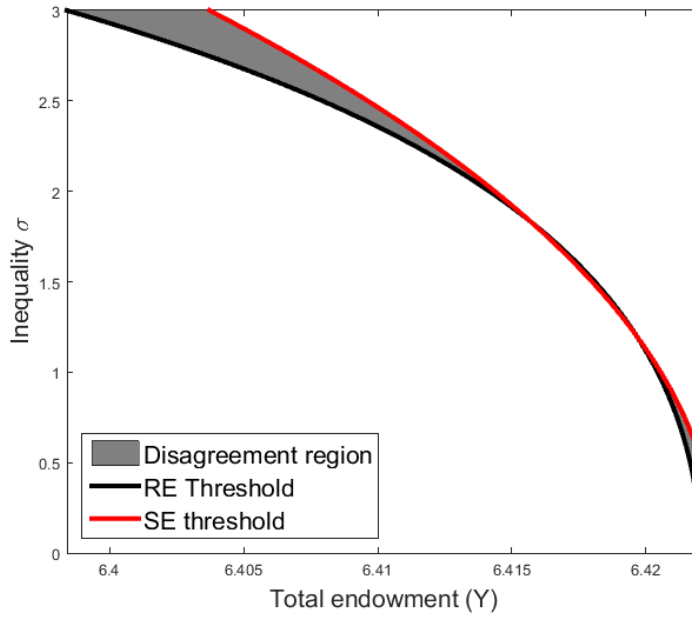
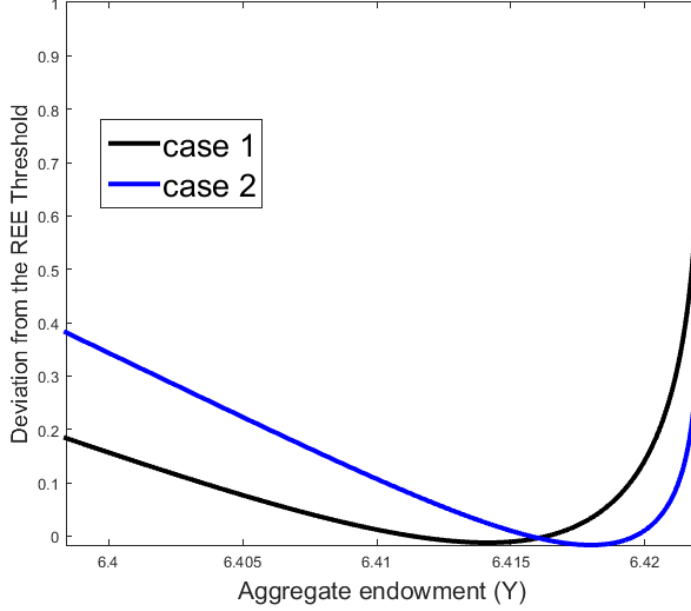
Figure 5: Subjective expectations based only in  $Y$ Figure 6: Self-Confirming Equilibrium expectations  $M(X_j, \beta_j^*)$ 



Figure 7: Deviation from the REE equilibrium



Case 1: Parametrized Expectations  
Case 2: Self-Confirming Expectations

along the equilibrium path induces a low deviation at high levels of  $Y$  and a high deviation for low levels, when firms do not tend to invest.

#### 4.7 Extension to a T-Periods Economy

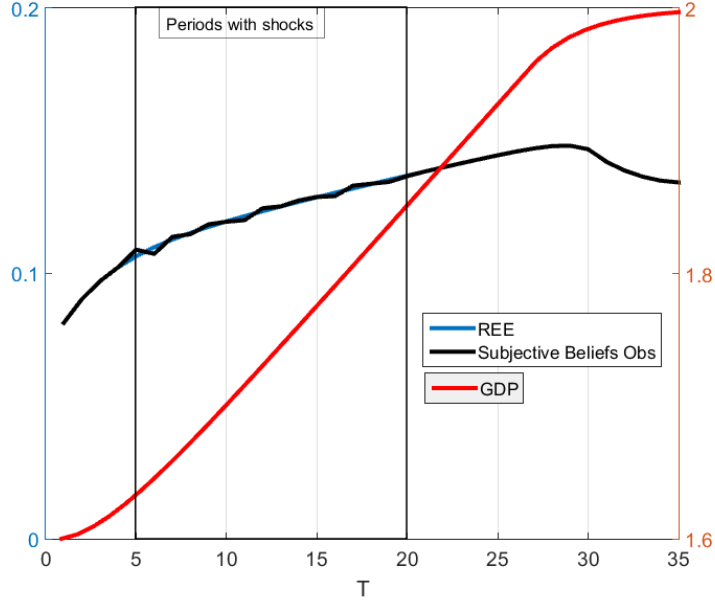
Here I extend the model to  $T$  periods to investigate the joint dynamics of  $Y$  and investment in the production of new qualities. The stochastic process for  $Y, \sigma$  is given by a vector autoregression specification. The economy starts with a lower level of average productivity with respect to the mean and I focus on the growth period during the convergent path.

In this version of the model firms are heterogenous on the innovation costs. These costs are calibrated to allow for a smooth entry of firms in the top quality level as the economy grows and converges to the long run mean of average productivity. Households have the same structure as before.

The path the economy follows when there are no shocks is shown in Figure 8. The Figure shows the path of GDP and investment at the REE and the SCE. Investment grows with GDP as firms invest to develop new goods. At the end of the time window shown GDP is already converging to the steady state level and investment slows down.

In this economy, I focus on the time span between periods 5 and 20 and allow shocks to the productivity distribution to change the path of the economy. Firms have to form expectations

Figure 8: Benchmark Scenario: Convergence to the steady state



about demand during this time span to decide if they invest to produce the high quality level  $q_h$  of their variety or not. I compute in this situation the rational expectations solution and the subjective beliefs of the SCE.

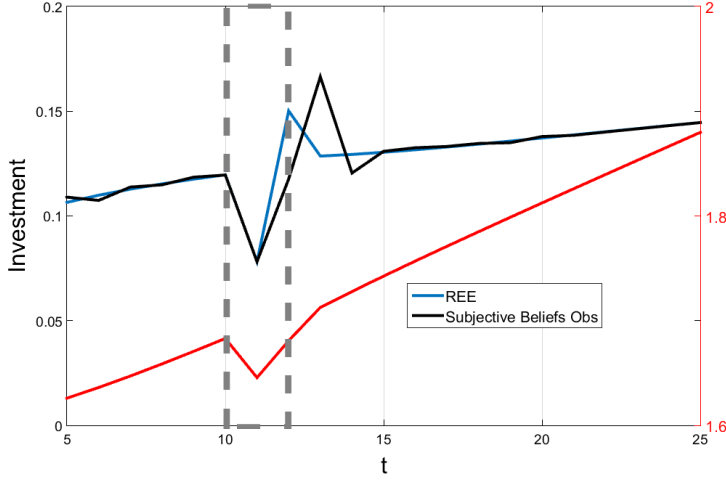
#### Impulse Response: A Short Lived Recession.

Figure 9 presents the impulse response of a negative shock to labor productivity  $Y$  that is accompanied with an increase in the dispersion of productivity  $\sigma$ . The shock hits the economy at period 11 and we see that  $Y$  recovers pre-recession levels after one period. We could consider it a recession that lasts 2 years.

Investment in the REE has a fast recovery and the year after the recession is at peak levels. On the other hand, in Figure 9 it can be seen that in the subjective beliefs case (SCE) the recovery of investment is delayed. The reason for this delay is pessimism by investors. In Figure 10 the subjective expectations at the SCE are presented and compared to the objective expectations. The expectations plotted are those of the firm  $j$  that decided to invest in the REE but delayed investment in the SCE.

First note that the period the shock hits (11) the economy we see a wedge on expectations. The subjective expectations are above the objective expectations on the impact, nevertheless this difference does not affect the equilibrium outcomes as in both cases firms optimal decision is not to

Figure 9: Impulse response: A negative productivity shock with an increase in inequality



invest. The subsequent period on the other hand, we have that pessimism arises and despite the wedge is smaller this time the equilibrium outcome is different. Some firms decide to invest in the REE and develop new goods while that decision is delayed in the SCE. In period 13 when the firm decides to invest the expectations are again aligned and close to each other. This is what I refer by saying that pessimism that the excess in pessimism is not revealed.

#### Impulse Response: A Short Lived Expansion.

Consider the case of shocks of the same magnitude but with opposite direction. Figure 11 shows the analogue graph for the case just discussed. In this case there is no major difference between the timing and level of the reaction of investment. If anything, there is a minor delay in investment a few periods after the shock with the subjective expectations.

This result should come at no surprise. Firms forecasting scheme is quite precise during expansions as in this situations is when firms tend to invest and observe the outcomes or projects. Pessimism can be sustained along the equilibrium path during recessions while "optimism" is not a feature that the mechanism proposed generates during expansions. The intuition is simple as its the key argument in my analysis. Optimism is likely to be discovered because induces firms to invest while pessimism might be sustained as it is not confronted with observed outcomes. This difference is highlighted in Fudenberg and Levine (2009). The authors associate optimism with the Lucas critique; wrong beliefs that induce to take an action that reveals the mistake. The opposite case is pessimism and the SCE were wrong beliefs are not confronted with experience as the payoff is not observed.

Figure 10: Comparison Between Subjective and Objective Expectations

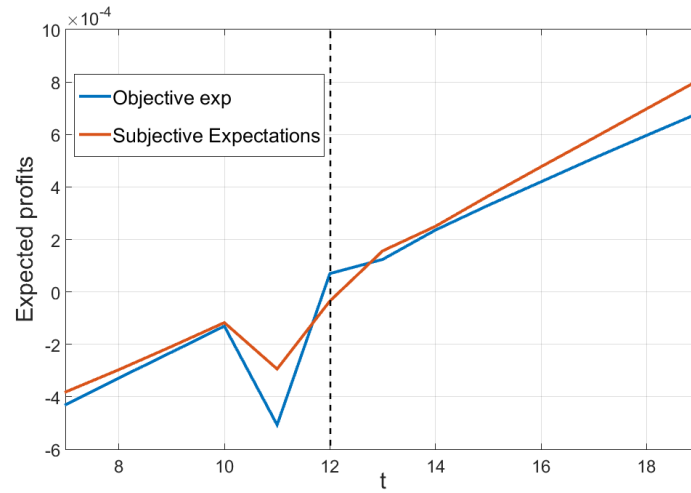
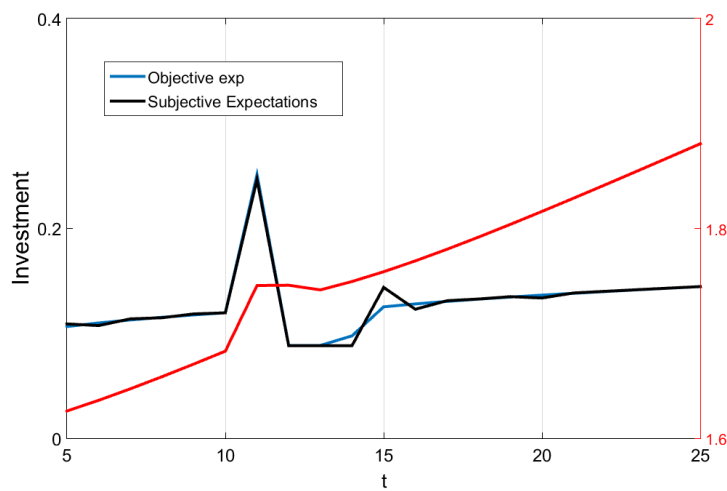


Figure 11: Impulse response: A positive productivity shock with a decrease in inequality



## 5 Full Model

The delay in investment is inefficient and only generated by the imperfect market knowledge that firms have. This open the following questions: 1) In which type of markets is it more likely to find a pessimistic delay? 2) For which type of recessions? and 3) How can policy be designed to ameliorate the welfare losses? The simple model has features that does not allow us to study these questions as innovation is only possible at the highest quality and the income process and the opportunity cost of investment are exogenous. Next, I extend the model to be able to address these questions.

### 5.1 Setup

The extension of the model is done in three dimensions

1. Firms can invest to innovate in three different quality levels: low quality; medium quality; and high quality.
2. Households decide savings and there is a continuous distribution of labor productivity.
3. The basic good is produced with capital and labor. Capital is held by the differentiated goods firms or the households.

As before I focus on the transition path of the economy to an steady state with no growth and allow for shocks to arrive during the transition. The shocks considered in this setup are a productivity shock and a financial shock. The productivity shock is a change in the distribution of labor productivity and the financial shock is an increase in the cost of funding for the firms.

#### Households

Households supply inelastically one unit of labor. The household decides how much to buy of the asset ( $b_t$ ), expenditures  $X$ , and which differentiated good to buy on the market. The income sources are labor income, the return on asset holding and the dividends payed by the differentiated good firms.

The households have to decide the level of expenditures at period  $t$  before the taste shocks are realized. This way all households with the same productivity and assets will decide the same level of expenditures although they might decide to buy a different bundle. The problem of the household

is to

$$\begin{aligned}
\max \quad & E \left\{ \sum_t \beta u^i(z, j, q) \right\} \\
s.t \quad & u^i(z, j, q) = zq + \varepsilon_{j,q}^i \\
& X = wa_i + (r_t + p_t)b_{t-1} - p_tb_t + div_t^i \\
& X \geq z + p_{j,q}
\end{aligned} \tag{14}$$

where  $div_t^i$  are the dividends paid to household  $i$ ,  $w$  is the wage,  $a_i$  is the productivity and  $r_t$  is the net return on the asset holdings  $b_{t-1}$  and  $p_t$  is the price of the asset at  $t$ . Expenditures are spent on the basic good  $z$  (numeraire) and the unit of the differentiated good, paid at price  $p_{j,q}$ . Constraint 14 is similar to the constraints used in the cash in advance literature. The household has to decide in the "morning" how much resources to take to the goods market that opens in the "afternoon". The taste shock is realized upon arrival to the goods market.

First I have to characterize the expected instantaneous utility of expenditures  $X$ . Following Verboven (1996) we have that the expected utility of a level of expenditures  $X$  is given by:

$$E \{ u^i(z, j, q) \mid X \} = \ln \sum_{\omega \in Q} \left[ \sum_{j \in \mathbb{J}_\omega} e^{(X - p_j)\omega / \theta_\omega} \right]^{\theta_\omega}$$

then the expected marginal utility is

$$\frac{\partial E \{ u^i(z, j, q) \mid X \}}{\partial X} = \frac{\sum_{\omega \in Q} \omega \left[ \sum_{j \in \mathbb{J}_\omega} e^{(X - p_j)\omega / \theta_\omega} \right]^{\theta_\omega}}{\sum_{\omega \in Q} \left[ \sum_{j \in \mathbb{J}_\omega} e^{(X - p_j)\omega / \theta_\omega} \right]^{\theta_\omega}}$$

and the corresponding Euler equation corresponds to

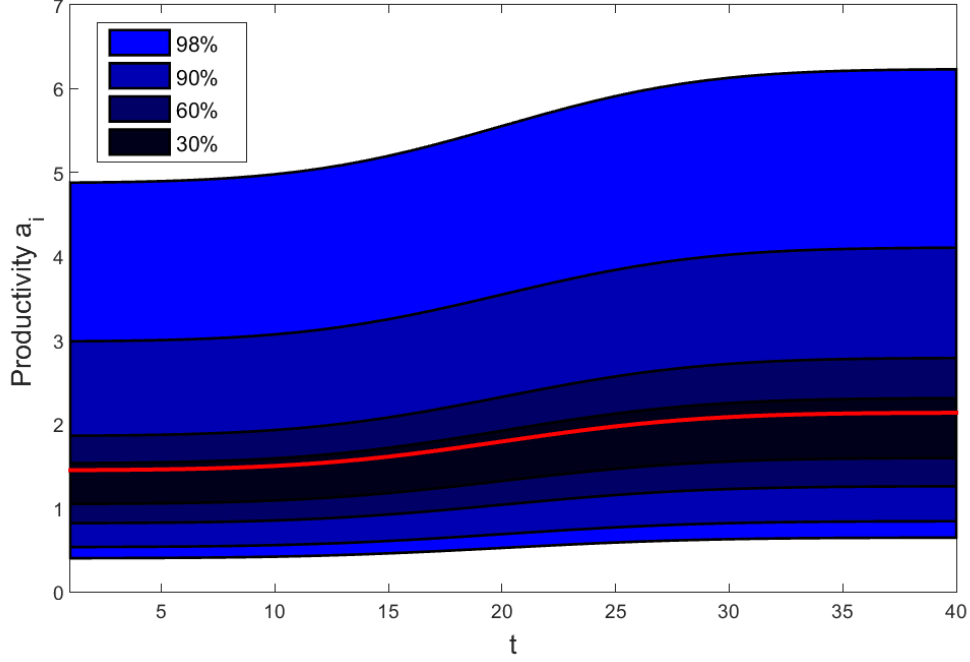
$$\frac{\frac{\partial E_t \{ u_t^i(z, j, q) \mid X \}}{\partial X}}{E \left\{ \frac{\partial E_{t+1} \{ u_{t+1}^i(z, j, q) \mid X \}}{\partial X} (p_{t+1} + r_{t+1}) \right\}} = \frac{\beta}{p_t}$$

For simplicity and tractability I impose that the agent has a policy function that determines savings that is linear mapping of time, previous asset held and labor productivity. The mapping is selected such as the savings decision minimizes the deviations from the Euler equation.

### A More General Endowment Distribution

At any point in time the skills distribution is characterized by a Beta distribution for the lowest 85% of the productivity distribution and a Pareto distribution for the upper tail. Absent of shocks labor productivity grows during the transition period. Growth is modeled by an increase in the limits

Figure 12: Skills distribution over time



of the support of the Beta distribution. Without shocks the productivity distribution follows the path in Figure 12. Along the transition and also when we allow for shocks the ordering of agents along the distribution productivity is not altered. This means that we allow for aggregate and idiosyncratic income risk but not for any shift in the order of agents with respect to productivity.

### Multiple Quality Levels

Each firm in the differentiated goods market has two quality level of their variety in the market and the investment decision resides on the possibility to develop a higher quality level. Compared to the basic model the key modification I introduce here is that not all firms have the same two outstanding qualities. In total there are 5 possible quality levels in the economy  $\{q_0, q_1, q_L, q_M, q_H\}$ . The firms are divided into three sets:

1. Low quality firms: firms with outstanding qualities  $q_0$  and  $q_1$  that can innovate to generate  $q_L$ .
2. Medium quality firms: firms with outstanding qualities  $q_1$  and  $q_L$  that can innovate to generate  $q_M$ .
3. High quality firms: firms with outstanding qualities  $q_L$  and  $q_M$  that can innovate to generate  $q_H$ .

### Differentiated goods firms

Firms have Self-confirming expectations with a forecasting function of profits  $M(X_j; \beta_j)$ . Firms also face technological risk: There is a probability  $p_f$  that if the firm invests there the new good is not successfully developed.

The firm can finance the investment to develop a new quality with retained earnings  $re_{j,t}$  or with a loan going short on the asset  $b$  in the economy. The opportunity cost of retained earnings is the return on the asset  $b$  given by  $\frac{(p_{t+1} + r_{t+1})}{p_t}$ . On the other hand, if the firm decides to use a loan it has to pay a premia given by  $\varrho_t$ . Since there is no default in this setup and all firms can eventually repay the debt we have that  $\varrho_t$  reflects a real cost on the reallocation of resources in the economy. I set this cost to be 0 if no shocks hit the economy. The financial shock will affect this term, increasing the cost on the reallocation of resources.

The optimal investment rule of the firm is

$$M(X_j, t; \beta_j)(1 - p_f) \geq E_t \left\{ \bar{f}_{q,t} \frac{(p_{t+1} + r_{t+1})}{p_t} + (\bar{f}_q - re_{j,t}) \varrho_t^1 \right\}$$

where the left hand side has the expected profits and the right hand side the cost. Expectations about the interest rate are the same for the whole economy and all agents have the same expectations. Note that  $(\bar{f}_q - re_{j,t}) \varrho_t$  represents the premia paid on borrowed resources.

The optimal pricing is given by:

$$p_q = c_q + \frac{\theta_q}{q}$$

and the profits of the differentiated goods firms correspond to:

$$\pi_j = \sum_{q \in Q_{j,t}} (D_{j,q}(\cdot) [p_q^j - c_q] - f_q)$$

where  $f_q$  is the fixed cost of variety  $q$ . I suppose that firms pay as dividends a fix fraction  $d$  of their total income after investment as given by:

$$div_{j,t} = d((\pi_t - \chi_{j,t}) + (r_t + p_t) re_{j,t-1})$$

and the top 15% households are the owners of the firms. Consequently the assets of the firm evolve as follows:

$$re_{j,t} = (1 - d) \left( \frac{(\pi_t - \chi_{j,t}) + (r_t + p_t) re_{j,t-1}}{p_t} \right)$$

where  $\chi_{j,t}$  is the investment of firm  $j$  at period  $t$  in the development of a new good.

### Mutual Fund

Similar to Gornemann, Kuester, and Nakajima (2012) households and firms save and borrow using a single asset  $b$ . This asset correspond to a share in a mutual fund that owns the capital of the



economy. This fund also decides how much capital to rent in order to maximize the return on its portfolio. Therefore, the fund set the capital utilization  $u_t$  to maximize the payments on capital.

Depreciation is suppose to depend on the level of utilization as follows

$$\delta(u_t) = \delta_0 + \delta_1(u_t - 1) + \frac{\delta_2}{2}(u_t - 1)^2$$

The fund behaves competitively and does not internalize the effect of its decision on  $u_t$  on the price of capital. Therefore the optimal utilization level solves

$$u_t = \frac{R_t/p_t - \delta_1}{\delta_2} + 1$$

where  $R_t$  is the rental rate of capital paid by firms. The price of an asset at the mutual fund is  $p_t$  that is equal to the price of capital. The dividend paid for each unit of capital to the asset holders of the mutual fund is given by

$$r_t = R_t - \delta(u_t) - \varrho_t^2$$

where  $\varrho_t^2$  is a real cost on the allocation of resources. In this case the lending of capital. At the steady state  $\varrho_t^2 = 0$ ; this is the other source of financial shocks.

### Basic Sector $z$ Firms

The representative firm hires all the labor and produces with the production function

$$Y_t = A_t^{1-\alpha} K_t^\alpha$$

where  $A_t = \int_i a_i \partial H(a_i)$ .

The problem of the firm is

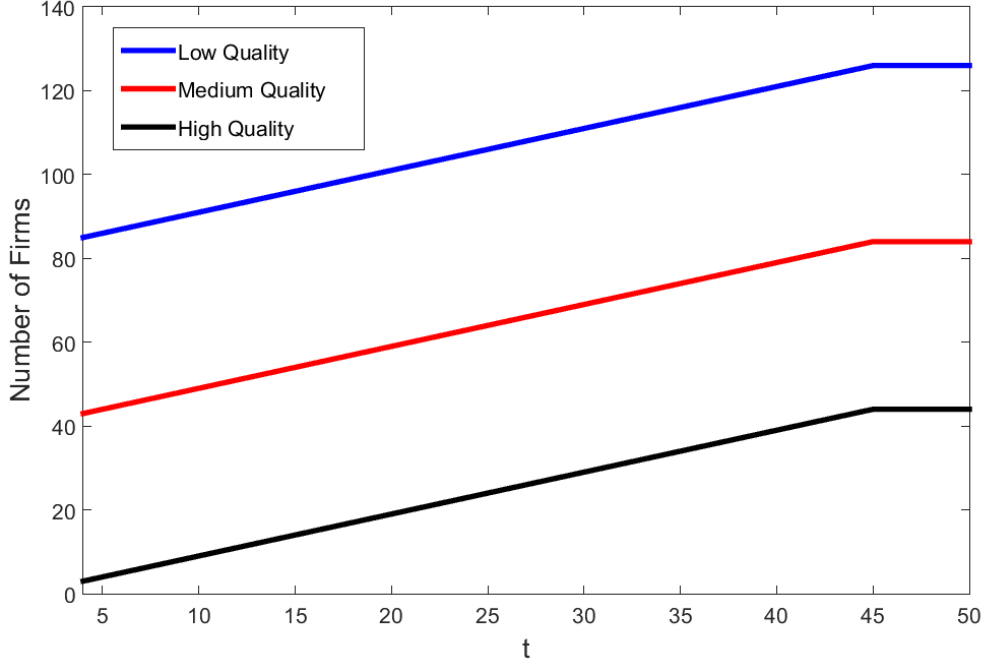
$$\max_{A_t, K_t} A_t^{1-\alpha} K_t^\alpha - \omega A_t - R_t K_t$$

the firm has constant returns to scale and therefore the scale is not determined. The equilibrium prices of inputs are given by

$$\begin{aligned} \omega_t &= (1 - \alpha) \frac{Y_t}{A_t} \\ R_t &= \alpha \frac{Y_t}{K_t} \end{aligned}$$

so output is demand driven and it determines wages and the rental rate of capital.

Figure 13: Firms Entry to each quality level



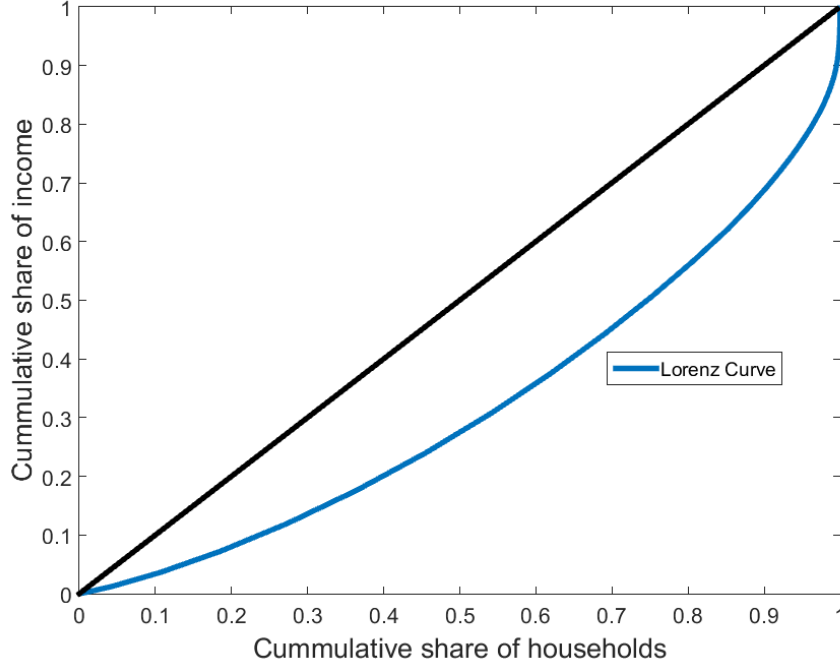
### The Productivity Shocks and the Financial Shock

The productivity shocks are two: A shock to the  $\beta$  of the Beta distribution of productivity and a shock to the upper limit of the support of the Beta distribution. Both shocks follow an auto-regressive process. The financial shock is constituted by an increase of  $\varrho_t^1$  and  $\varrho_t^2$ . Both terms generate a real cost in the allocation of resources.  $\varrho_t^1$  increases the cost of funding for the firms in the differentiated goods sector. Analogously  $\varrho_t^2$  increases the cost of resources for the basic goods firms. Both shocks follow an auto-regressive stationary process as well.

## 5.2 Equilibrium Dynamics and Impulse Responses

The fix cost for the development of a new quality level are calibrated to have a smooth entry of firms in the absence of shocks. Figure 13 shows the number of firms at each level that have developed the corresponding quality level for their variety at each period. This is the case where the economy faces no shocks. The Lorenz curve once the economy converges to the steady state is presented in Figure 14.

Figure 14: Lorenz Curve



### 5.2.1 Productivity Shock

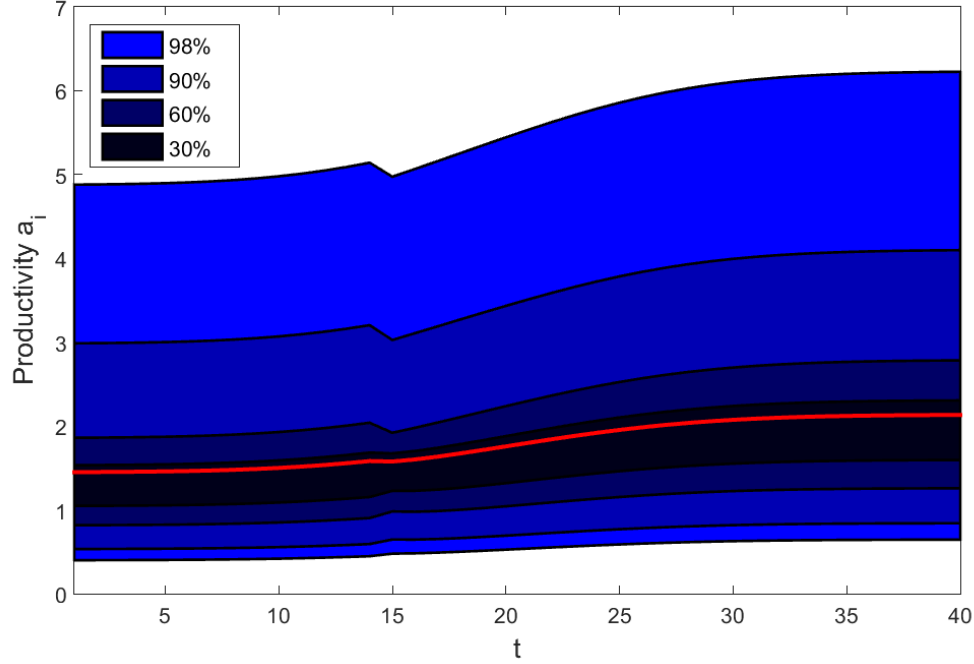
The shock affects the distribution of skills, It is a shock that reduces aggregate skills and at the same time reduces the dispersion of skills. The evolution of the productivity distribution with the shock hitting the economy in period 15 is shown in Figure 15.<sup>13</sup>

The dimension I will focus to assess the delay in investment is the number of firms that develop the new version of their differentiated variety. I compare the number of firms at each of the 3 sets of firms (Low, medium and high quality) that have invested to develop the higher quality in the SCE and compare it with the case with objective expectations.

Figure 16 shows the number of firms at each quality level of innovation. We see that in the three cases subsequent to the shock, the number of firms that develop the new quality is below the counterfactual with objective beliefs (dashed line). This pattern shows to be more persistent for the high quality good case although it is also clearly seen at the low level of quality.

<sup>13</sup>Depending how inequality is measured it can increase or reduce inequality. If it is measured by the variance of skills it is reduced. The shock increases the  $\beta$  of the beta distribution and reduces the upper limit of the support of the Beta distribution. The tail given by the Pareto distribution is shifted by the distribution stays the same and has the same mass of workers in all periods.

Figure 15: Productivity Shock



### 5.2.2 Financial shock

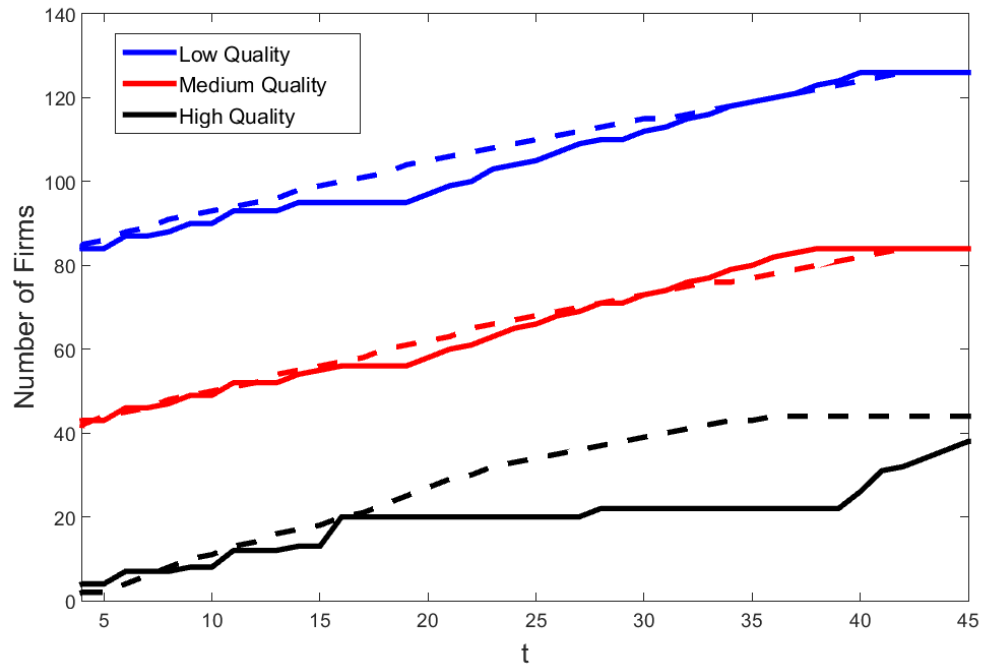
The financial shock is an increase in  $\varrho_t^1$  and  $\varrho_t^2$  that vanishes with time. In this case the cost for financing the investment is increased. The shock generates a wedge between the interest rate paid by credit and the return to asset holders; it is a real cost to the reallocation of resources. Figure 17 shows the number of firms at the three levels of innovation. Pessimism emerges primarily again for the high quality version of the goods.

The main difference to highlight with the productivity shock is the delay in the emergence of pessimism. The shock arrives also at period 15 and pessimism arises many periods afterwards even when the shock has vanished. The mechanism behind this result is that is not the increase in the cost of resources itself what drives pessimism. It is the induced change in the income distribution that is an outcome of the financial shock what changes the relative demands for goods in a way that is not fully acknowledged by the firms expectations.

The fact that pessimism emerges with a delay does not mean that the shock has not immediate consequences in the investment decisions. The investment decisions are indeed affected by the shock, just that both objective and subjective beliefs are aligned. This is because the shock is over the cost of investment and not over the structure of demand, the main source of uncertainty of firms when they decide if to innovate or not.

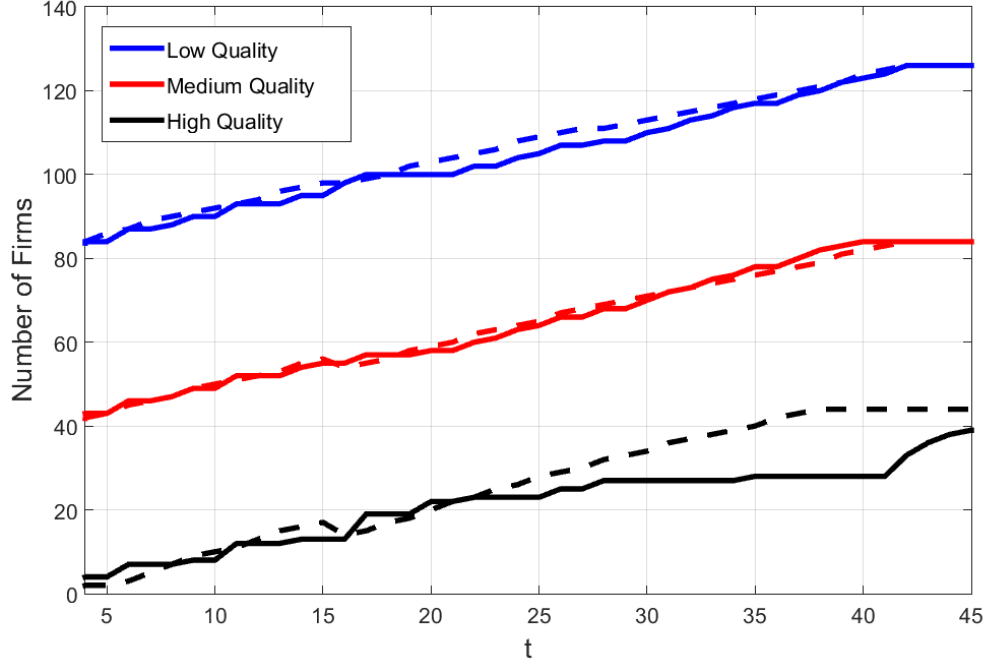
Finally in Figure 18 the GDP for the three cases is presented. The economy with no shocks, with

Figure 16: Firms Entry, Labor Productivity Shock.



Solid line: Entry at the Self-Confirmed Equilibrium  
Dashed line: Entry with objective Expectations

Figure 17: Firms Entry, Financial Shock



Solid line: Entry at the Self-Confirmed Equilibrium  
Dashed line: Entry with objective Expectations

the productivity shock and with the financial shock. The productivity shock generates a strong response in GDP with an immediate fall. On the other hand, the financial shock has a smaller effect but the recovery is slower and the convergence to slower. I now discuss policy alternatives to avoid a delay in investment and generate a faster recovery after the recession.

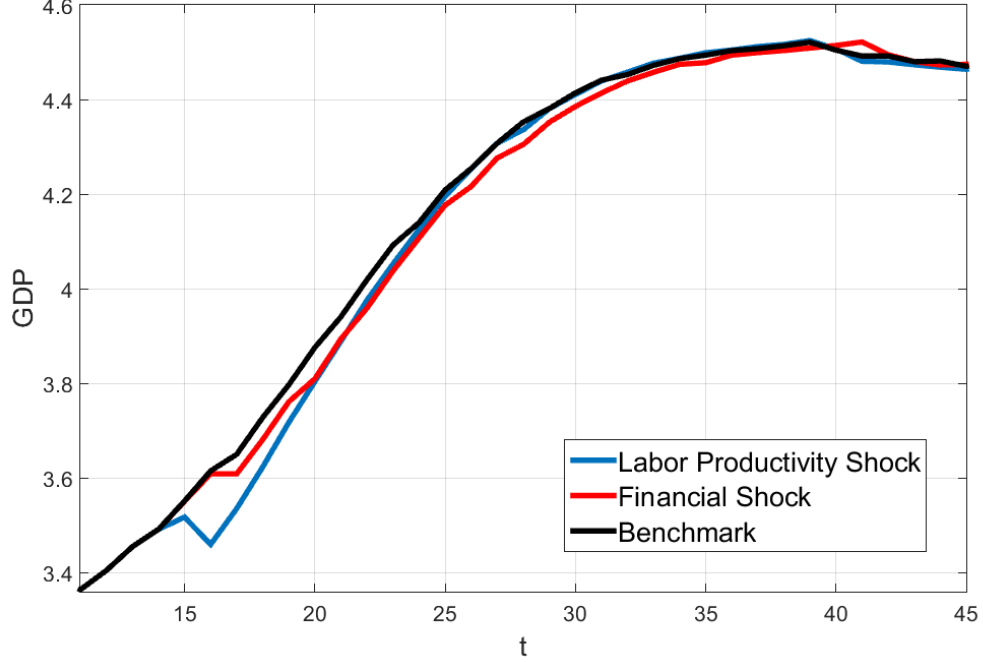
## 6 Policy interventions

In this section I analyze the role of policy to incentivize investment and overcome the pessimistic delay. First, I argue that a policy based on a countercyclical corporate income tax can be more effective to screen and target firms with pessimism vis-a-vis firms that do not invest because of high technological risk. Second I show the welfare gains of such a policy across the population when investment at different quality levels is targeted.

### 6.1 Corporate Income Tax and the screening of pessimism.

This section provides a preview on how policy can undo the pessimistic delay. The focus is how to provide incentives for the firm to carry on the investment. Going back to the optimal investment

Figure 18: GDP for the three cases: No shocks, Productivity shock and Financial shock.



decision of the firm, we have that a firm invests if

$$M(X_j, t; \beta_j)(1 - p_f) \geq \bar{f}_{q,t} E_t \left\{ \frac{(p_{t+1} + r_{t+1})}{p_t} \right\} + (\bar{f}_q - r e_{j,t}) \varrho_t^1$$

consider the case where the firm is pessimist, i.e the following condition is satisfied

$$M(X_j, t; \beta_j) < E \{ D_{j,q}(\cdot) \mid X_j \}$$

such that subjective beliefs are lower than objective. Furthermore, take the case where the firm does not invest because of pessimism, the case we are interested is where

$$M(X_j, t; \beta_j)(1 - p_f) < \bar{f}_{q,t} E_t \left\{ \frac{(p_{t+1} + r_{t+1})}{p_t} \right\} + (\bar{f}_q - r e_{j,t}) \varrho_t^1 < E \{ D_{j,q}(\cdot) \mid X_j \} (1 - p_f) \quad (15)$$

The first inequality in 15 establishes that the firm will not invest and the second inequality that it is optimal to invest given the information available. This is the particular case we would like to avoid with policy.

I will consider two policy interventions that can potentially provide the incentives for the firm to invest.

1. Provide subsidies to investment. Subsidize a fraction of the cost  $\bar{f}_{q,t}$ .

2. Corporate income taxation. Lower the tax on profits by  $T(\pi_t)$ .

If the policies can be targeted at the investing firm these two possibilities can induce the pessimist firm to invest. The first by lowering the cost of investment and the second by increasing the expected profits. Nevertheless, to implement this policy we suppose the government knows if the firm is being pessimistic or not. This imposes that the government has a greater deal of information and can fully spot when pessimism arises.

A firm might delay investment even if it is not pessimistic but because it faces a temporary increase in  $p_f$ : the probability the new good is not developed. This is a delay in investment generated by an increase in technological risk. There is a lower expected profits but this is an objective assessment of risk. I suppose that the government cannot observe  $\rho_f$  so it is observationally equivalent pessimism to technological risk, recall that for the expected profits both enter in the same way. In particular, the two recessions we considered could be accompanied by an increase in technological risk. First, A shock to productivity could also affect the technology of innovations. Second, the increase in the cost of funds could also be the result of higher risk in investments rather than a financial shock.

Consider two firms that have the same expected profit of investment. Firm  $j$  is pessimistic about demand and has low technological risk ( $p_f^j$ ) while firm  $m$  is not pessimistic but has high technological risk ( $p_f^m$ ). Since both have the same expected gain from the investment we have

$$M(X_j, t; \beta_j)(1 - p_f^j) = M(X_m, t; \beta_m)(1 - p_f^m)$$

and

$$\begin{aligned} M(X_j, t; \beta_j) &< M(X_m, t; \beta_m) = E\{D_{j,q}(\cdot) | X_j\} \\ p_f^j &< p_f^m \end{aligned}$$

Note that an investment subsidy that makes it optimal for the firm  $j$  to invest also makes it optimal for  $m$  to invest. Despite the fact that the benefit for  $j$  investment is greater than that of  $m$  because of the lower risk. On the other hand, a lower marginal tax on profits that makes the firm  $j$  indifferent to invest or not assures that

$$(1 - p_f^j)T(M(X_j, t; \beta_j)) + M(X_j, t; \beta_j)(1 - p_f) = \bar{f}_{q,t}E_t\left\{\frac{(p_{t+1} + r_{t+1})}{p_t}\right\} + (\bar{f}_q - re_{j,t})q_t^1$$

where  $(1 - p_f^j)T(M(X_j, t; \beta_j))$  is the expected tax deduction. Note the effect is given by the tax discount multiplied by  $(1 - p_f^j)$ . For the firm  $m$  the tax incentive is diminished by the higher technological risk (lower  $(1 - p_f^j)$ ) and as long as the discount does not grow with profits the firm  $m$  will not invest.

The corporate income tax can allow the planner to screen firms between technological risks and expectations about demand. Next we evaluate the welfare gains of the alternative policy interventions for the two sources of investment delay; technological risk or pessimism.



## 6.2 Comparison of policies

Now I proceed to compute the welfare gains of the two types of policies: An investment subsidy and a lower corporate income tax. For both policies I set the scale of the intervention to be the minimum possible that provides incentives for an additional firm to invest and develop a new good. The policy is targeted at a given quality level and I consider the three cases at hand: low quality, medium quality and high quality firms.

I will focus on the recession driven by the technological shock as presented in section 5.2.1. In this case we have a delay in investment driven by pessimism at the three quality levels. For each quality level I evaluate the welfare gains along the population of the two policies and decompose it into its key drivers.

### Investment Subsidy

The investment subsidy corresponds to a lump-sum transfer given by the government to the firm that invests to produce a new quality. It is a fraction  $\zeta_t$  of the investment cost  $\bar{f}_{q,t}$  and is targeted to a particular quality level.

This policy affects directly the cost of the investment for the firm. The size of  $\zeta_t$  is set to the minimum level necessary to have that the firm that an additional firm invests in period  $t$ . Formally, following from the optimal condition for investment, it follows that  $\zeta_t$  is given by the solution to:

$$M(X_{j^*}, t; \beta_{j^*})(1 - p_f) = (1 - \zeta_t) \bar{f}_{q,t} E_t \left\{ \frac{(p_{t+1} + r_{t+1})}{p_t} \right\} + ((1 - \zeta_t) \bar{f}_q - re_{j^*,t}) \varrho_t^1$$

this equation just states that with the investment subsidy the firm  $j^*$  is indifferent between investing or not. The firm  $j^*$  is set to be the firm that requires the smallest subsidy to engage in investment.

The purpose of setting the intervention at this level is to measure the welfare gains of pushing an additional firm to invest using policy. Taking into account that any resource allocated to such firm has to be collected with taxes and also that it has the corresponding general equilibrium effects. It is somehow the “marginal” effect of policy evaluated at a situation with no policy.

### Corporate Income Tax

The corporate income tax policy intervention consists on a lower corporate tax to  $(1 - \tau_t^c)$  for the next periods after the policy is announced. The deductions continue up to when a maximum level of total subsidy  $\bar{S}_t$  in present value is reached. Therefore, the firm that is benefited with the policy can expect the tax deduction as long as the total subsidy cap  $\bar{S}_t$  is not reached.

The profits subsidy is granted only for the new created quality and targeted again at the “marginal” firm. Just as the investment subsidy. Therefore, if a firm invests to develop a new good but the investment fails then the firm cannot claim the tax discount as there are no profits coming from the production of the new good.

$\bar{S}_t$  is set to the minimum level that provides the necessary incentives for an additional firm to invest in period  $t$ . Formally it is given by the solution to:

$$(M(X_{j^*}, t; \beta_{j^*}) + \bar{S}_t)(1 - p_f) = \bar{f}_{q,t} E_t \left\{ \frac{(p_{t+1} + r_{t+1})}{p_t} \right\} + (\bar{f}_q - re_{j^*,t}) \varrho_t^1$$

Note the key difference between the two policies, that was already highlighted in the previous section. While the investment subsidy depends on the product  $M(X_{j^*}, t; \beta_{j^*})(1 - p_f)$ , the corporate tax policy is affected differently by each component. The Technological risk and the expectations of future profits have a different effect on the scale needed to push an additional firm to invest. The corporate income tax allows to screen firms by their technological risk. Therefore, the policy can be set to be effective up to a maximum level of technological risk  $\rho_f$  and set  $\bar{S}_t$  accordingly. If the firms have a higher technological risk then the policy will not provide enough incentives for the firms to invest.

### Financing and Timing of Policies

The government follows a balanced budget every period and therefore every subsidy is backed with taxes on labor. Taxes are collected with a flat rate  $\tau_{w,t}$  at each period  $t$  depending on the needs of the government that period to cover the cost of the subsidies. The policies to break the delay of investment are implemented the first period after the economy presents positive growth after a recession (negative growth rates). It is a simple rule to time the beginning of the recovery.

### Welfare effects decomposition

The policy interventions described affect the households of the economy through different channels. I decompose the total welfare effect of the policy into four components as follows:

1. **Variety expansion:** When a firm develops a new good the varieties available for consumers expand and consequently the utility level they can reach with a given level of income.
2. **Wage change:** The wage is valued in units of the numeraire good: the basic good. Once resources are devoted to the production of a new variety this implies a lower investment to accumulate capital in the basic sector and therefore has effects on wages.
3. **Labor Tax change:** The burden of the policy intervention lies on the workers. This channel measures the cost of the tax change on the welfare of each household.
4. **Capital Income change:** The decision to invest by a firm affects the level of capital accumulated in the basic good sector (The alternative investment opportunity) and also the dividends paid by the firm that invests and all of its competitors. The capital income change measures the change in the return of capital in the basic good sector and the change in dividends paid in the differentiated goods sector.

The policy interventions that induce an additional firm to invest have a unambiguous positive effect on the first component and negative on the second and the third. The fourth component can go both ways depending on the portfolio of the household. Households gain by the higher return on capital on the basic sector and might be worse off by the change in dividends paid in the differentiated sector.

Notice that while for a single firm the decision to invest is optimal and maximizes the present discounted value of distributed dividends, the firm does not take into account the lower profits that it generates for its competitors. So it is the higher competition that might marginally reduce overall profits, what can generate that households that mostly held the differentiated firms assets can be negatively affected by the policy. This should come as no surprise as the government is subsidizing entry to a monopolistic market and this reducing monopolistic rents.

### 6.3 Results for the Recession Caused by the Productivity Shock

Figures 19, 20 and 21 show the decomposition of the welfare gains for the corporate income tax policy for the recession generated by a productivity shock. The decomposition is done for each household in the economy, where the index of the household  $i \in [0, 1]$  corresponds to its ranking in the productivity distribution.

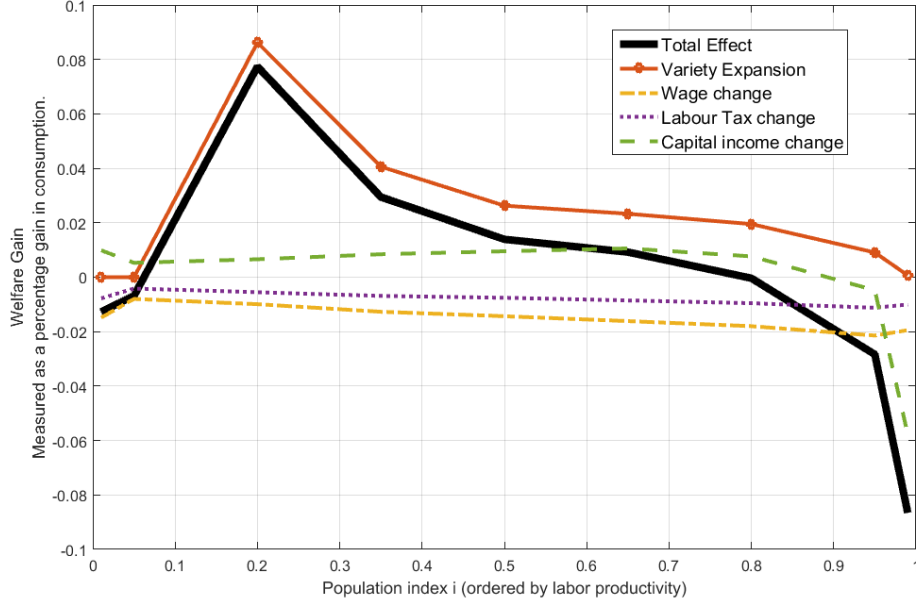
Each of the graphs present the welfare gains for the policy targeted at the three quality levels. Figure 19 shows the welfare gains when the policy is targeted to the low quality firms. We can see that the expansion in varieties mostly benefits the households with low productivity and corresponding low income.

In all cases the greatest losses come for the wealthiest households, the reason is the lower distributed dividends. Also, the poorest households tend to be worse off if the policy is targeted at higher quality levels: low productivity households do not consume the newly created goods but suffer the loss in wages and raise in taxes (see Figure 21).

There is one additional effect of policy that can emerge in this setup. The policy to incentivize an additional firm to invest at a given quality level will itself affect the demand for the other firms and potentially their decisions in the following periods. This can generate a “Domino” effect that can amplify the effects of policy as well as lower investment at the quality levels not targeted by the policy. A case where both of this situations take place is when the policy is targeted at the high quality level as represented in Figure 21. The expanded variety effect on welfare is positive and high for high productivity households, partly driven by subsequent entrants into the high quality fringe over the following periods after the policy is implemented. On the other hand, the poorer households face a negative sign on the effect of the expanded variety factor. This comes from the marginally lower entry at low quality levels triggered by the policy targeting the high quality goods.

The high quality case is qualitatively and quantitatively different to the other cases because of two reasons. First, the size of the investment is higher at that level and consequently implies a higher

Figure 19: Welfare Gains: Lower Corporate Income Tax Policy targeted at a “Low quality” firm.



reallocation of resources. Second, There are less firms competing at that quality level and therefore entry of an additional firm has a higher effect of profits and relative demands.

Overall, for an utilitarian planner the policy would be optimal in all three cases. Furthermore, if the planner can observe the individual productivities then a system of transfers allows the policy to be Pareto improving. This since the total gains by the households that are better off with the policy, measured in units of the basic good, are greater than the losses of the household that are worse off with the policy.

These figures are almost identical if the policy implemented is the investment subsidy and the same conclusions go through. Nevertheless, if we have that the source of the delay in investment is technological risk the corporate income tax policy has no effect on the economy while the investment subsidy still provides incentives for an additional firm to invest. In the next section I show that in that case such an intervention might not be welfare improving.

## 6.4 Delay in Investment Driven by Technological Risk

If the source of the delay in investment is technological risk (an increase in  $(p_f)$ ) rather than pessimism then the delay is not inefficient. In this case a policy that induces additional investment might not be welfare improving with a utilitarian social welfare function.

For the technological risk case I focus on the investment subsidy policy. Figures 22, 23 and 24 present the welfare gains for the policy targeted at the three quality levels.

Figure 20: Welfare Gains: Lower Corporate Income Tax Policy targeted at a “Medium quality” firm.

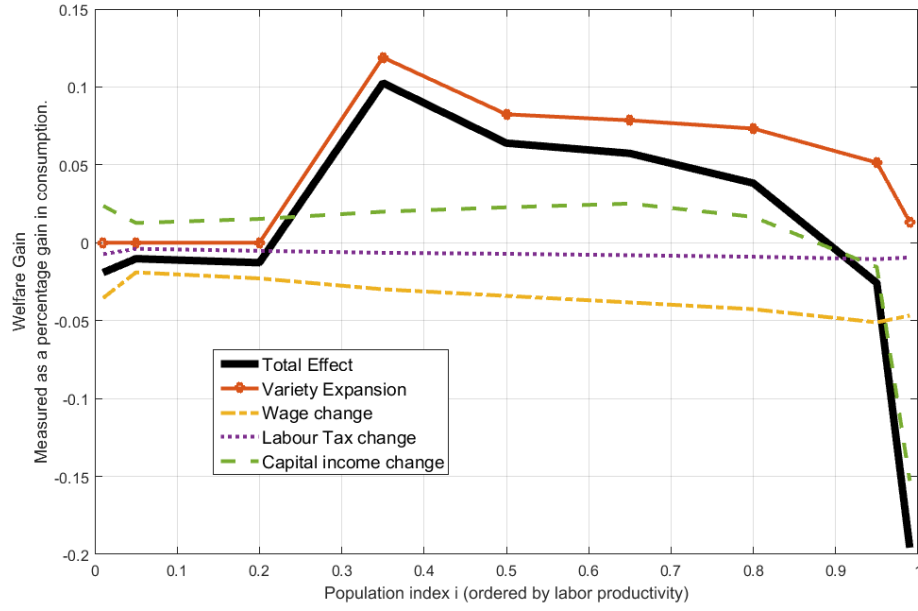
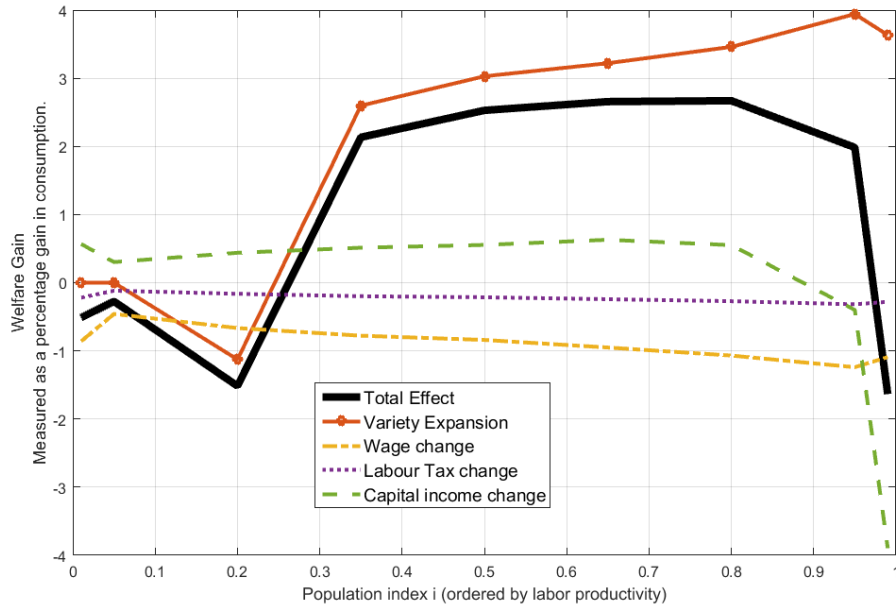


Figure 21: Welfare Gains: Lower Corporate Income Tax Policy targeted at a “High quality” firm.



The level of technological risk ( $p_f$ ) is set to be observationally equivalent to the pessimism case that was studied in the previous section. We observe a large decline in the welfare gains of the policy when the source of delay is technological risk. Furthermore, there is no system of transfers that would make the policy Pareto optimal in this situation.

The government is subsidizing an investment with high risk and generates excessive risk taking by firms in this economy. The risk is excessive given that the expected gains of the project are not large enough to compensate for the costs the economy is bearing by accumulating less capital and having lower wages. The higher probability of wasted resources by a failed investment project, where a firm does not fully internalize the costs of it, is the driver for the bad performance of this policy.

Therefore, if the government is uncertain about the source of the delay in investment a robust policy that would not trigger excessive risk taking would be to use corporate income taxation instead of investment subsidies. This prescription comes with a caveat, we focused on the recession driven by a shock to the productivity distribution. In the recession driven by the financial shock an investment subsidy has the additional advantage of relaxing the need of the firm for external funding.

Figure 22: Welfare Gains of Policy.  
Technological risk case, investment subsidy policy.  
Policy targeted at a “Low quality” firm.

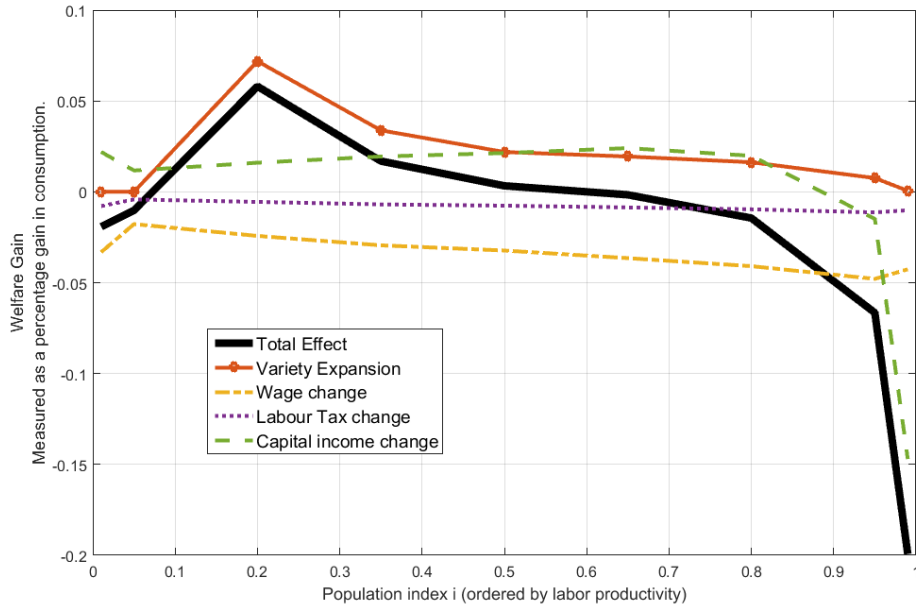


Figure 23: Welfare Gains of Policy.  
 Technological risk case, investment subsidy policy.  
 Policy targeted at a “Medium quality” firm.

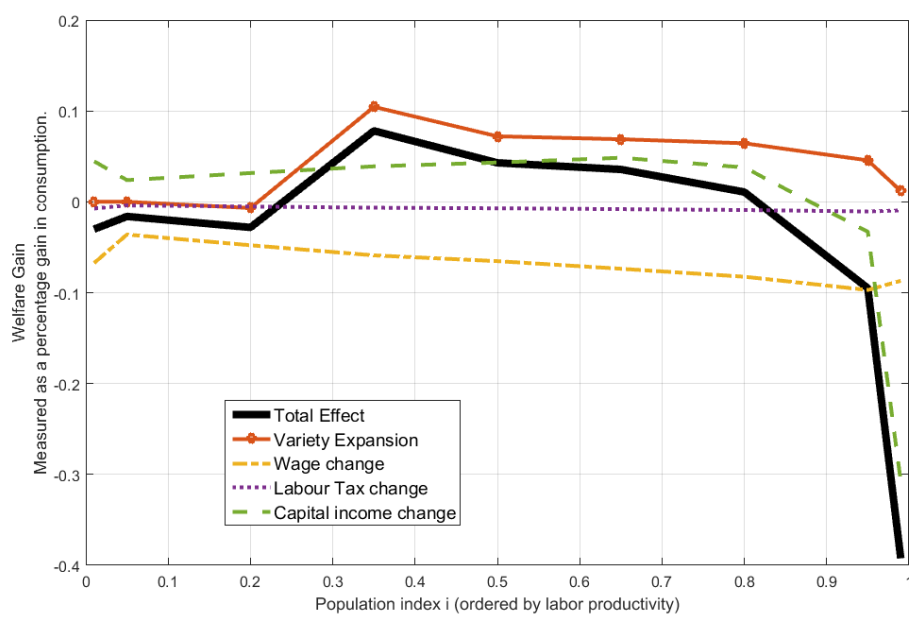
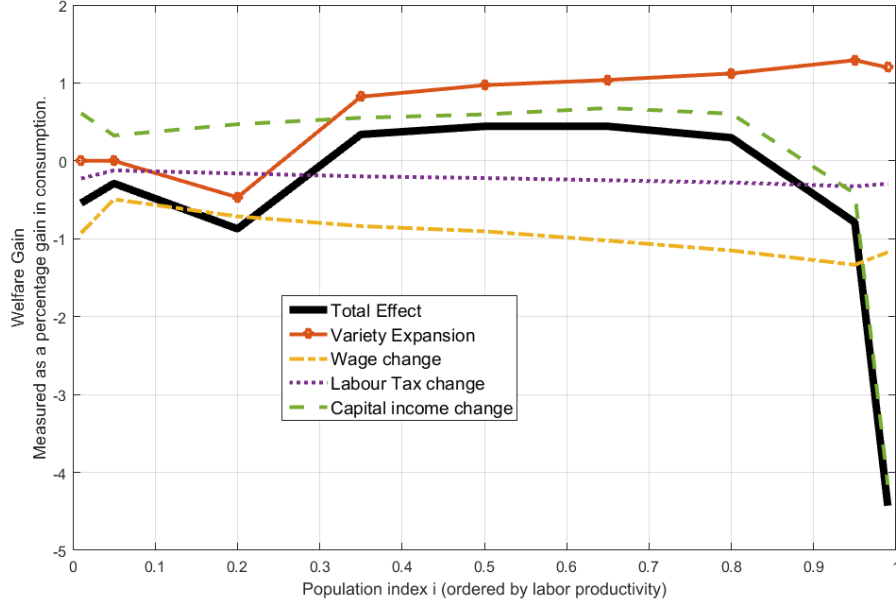


Figure 24: Welfare Gains of Policy.  
 Technological risk case, investment subsidy policy.  
 Policy targeted at a “High quality” firm.



## 7 Conclusions and Caveats

A stark prediction of macroeconomic models that study the business cycle is that after a recession investment tends to recover faster than GDP. In particular, during a recession, firms are able to identify emerging investment opportunities and thus investment leads the recovery. Contrary to this I have documented that investment recovers sluggishly after recessions and that GDP tends to lead the recoveries. I have put forward the idea that investors’ pessimistic beliefs contribute to this pattern.

The channel I propose is not suitable to be studied when the only equilibrium concept considered is the rational expectations equilibrium, as this restricts subjective beliefs to coincide with objective beliefs and thus discards pessimism. An alternative approach is therefore presented here. I study an economy where firms’ beliefs are simple forecasting schemes. Expectations are rational in the sense that the forecasting schemes are consistent with the probability distribution of equilibrium outcomes.

I have shown that during recessions investors might be excessively pessimistic about consumer demand and delay the implementation of projects. Furthermore, once they decide to invest, their excess in pessimism is not revealed and a SCE is obtained. Importantly, cyclical movements of the income distribution can affect the demand structures for the different goods available in the



economy and consequently the firms' investment. Economic models that maintain the assumption of homothetic preferences miss this dimension of variability, which in this paper is shown to have an important role to generate the delay in investment.

Taking this setting to the design of countercyclical policy, the policymaker faces the challenge of detecting the emergence of pessimism. I therefore study policy design when the government has uncertainty about the source of the delay in investment. Corporate income taxation might be a desirable instrument to use, as it is linked to the expected gains of firms, while interest rate policy or investment subsidies affect the cost of investment.

The theory needs further development in several dimensions. First, the stylized fact proposed needs to be further characterized, decomposing investment into the extensive and the intensive margin, also on the role of intangible assets. Second, how robust are the findings to the parameter values chosen and the assumption that firms model is linear. Third, the model assumes an extreme case where firms do not use the outcomes of competitors in their belief system. Finally an assessment of the quantitative relevance of the mechanism proposed.

## References

- ANDERSON, R. M., AND H. SONNENSCHN (1985): "Rational Expectations Equilibrium with Econometric Models," *Review of Economic Studies*, 52(3), 359–69.
- BLOOM, N. (2009): "The Impact of Uncertainty Shocks," *Econometrica*, 77(3), 623–685.
- DEN HAAN, W. J., AND A. MARCET (1990): "Solving the Stochastic Growth Model by Parameterizing Expectations," *Journal of Business & Economic Statistics*, 8(1), 31–34.
- FAJGELBAUM, P., G. M. GROSSMAN, AND E. HELPMAN (2011): "Income Distribution, Product Quality, and International Trade," *Journal of Political Economy*, 119(4), 721 – 765.
- FAJGELBAUM, P., E. SCHAAL, AND M. TASCHEREAU-DUMOUCHEL (2014): "Uncertainty Traps," NBER Working Papers 19973, National Bureau of Economic Research, Inc.
- FUDENBERG, D., AND D. K. LEVINE (1993): "Self-Confirming Equilibrium," *Econometrica*, 61(3), 523–45.
- FUDENBERG, D., AND D. K. LEVINE (2009): "Self-confirming equilibrium and the Lucas critique," *Journal of Economic Theory*, 144(6), 2354–2371.
- GABALLO, G., AND R. MARIMON (2015): "Breaking the Spell with Credit-Easing. Self-Confirming Credit Crises in Competitive Search Economies.," .
- GORNEMANN, N., K. KUESTER, AND M. NAKAJIMA (2012): "Monetary policy with heterogeneous agents," Working Papers 12-21, Federal Reserve Bank of Philadelphia.

## REFERENCES

## REFERENCES

- KHAN, A., AND J. K. THOMAS (2013): “Credit Shocks and Aggregate Fluctuations in an Economy with Production Heterogeneity,” *Journal of Political Economy*, 121(6), 1055 – 1107.
- KING, R. G., AND S. T. REBELO (2000): “Resuscitating Real Business Cycles,” NBER Working Papers 7534, National Bureau of Economic Research, Inc.
- KIYOTAKI, N. (1988): “Multiple Expectational Equilibrium Under Monopolistic Competition,” *The Quarterly Journal of Economics*, 103(4).
- KRUSELL, P., A. A. SMITH, AND JR. (1998): “Income and Wealth Heterogeneity in the Macroeconomy,” *Journal of Political Economy*, 106(5), 867–896.
- SARGENT, T. (2001): *The Conquest of American Inflation*. Princeton University Press.
- VAN NIEUWERBURGH, S., AND L. VELDKAMP (2006): “Learning asymmetries in real business cycles,” *Journal of Monetary Economics*, 53(4), 753–772.
- VERBOVEN, F. (1996): “The nested logit model and representative consumer theory,” *Economics Letters*, 50(1), 57–63.

## Chapter 3

# Optimal Redistribution with a Shadow Economy

[Joint work with Paweł Doligalski]

### Abstract

We examine the constrained efficient allocations in the Mirrlees (1971) model with an informal sector. There are two labor markets: formal and informal. The planner observes only income from the formal market. We show that the shadow economy can be welfare improving through two channels. It can be used as a *shelter against tax distortions*, raising the efficiency of labor supply, and as a *screening device*, benefiting redistribution. We calibrate the model to Colombia, where 58% of workers are employed informally. The optimal share of shadow workers is close to 22% for the Rawlsian planner and less than 1% for the Utilitarian planner. The optimal tax schedule is very different then the one implied by the Mirrlees (1971) model without the informal sector.

# 1 Introduction

Informal activity, defined broadly as any endeavor which is not necessarily illegal but evades taxation, accounts for a large fraction of economic activity in both developing and developed economies. According to Jutting, Laiglesia, et al. (2009) more than half of the jobs in the non-agricultural sector worldwide can be considered informal. Schneider, Buehn, and Montenegro (2011) estimate the share of informal production in the GDP of high income OECD countries in the years 1999-2007 as 13.5%. Given this evidence, the informal sector should be considered in the design of fiscal policy. This paper extends the theory of the optimal redistributive taxation by Mirrlees (1971) to the economies with an informal labor market.

The ability of the state to redistribute income depends on how responsive to taxes individuals are. When incomes are very elastic, differential taxation of different individuals is hard, because workers adjust their earnings to minimize the tax burden.<sup>1</sup> The shadow economy allows workers to earn additional income which is unobserved by the government. Without shadow economy, workers respond to taxes only by changing their total labor supply. With the shadow economy, they can additionally shift labor between the formal and the informal sector, which increases the elasticity of their formal income. As incomes in the formal economy become more elastic, redistribution becomes more difficult.

We show that the government can exploit differences in informal productivity between workers to improve redistribution. Suppose there are two types of workers: skilled and unskilled. The responsiveness of the skilled workers determines the taxes they pay and the transfers the unskilled receive. In the world without the shadow economy, this responsiveness depends on how easy it is for the skilled to reduce income to the level of the unskilled worker. If that happens, the government cannot tax differentially the two types of individuals. In the world with the shadow economy, the government can improve redistribution in the following way. By increasing taxes at low levels of formal income, the unskilled workers are pushed to informality. If the unskilled workers can easily find a good informal job, this transition will not hurt them much. Now the skilled workers can avoid taxes only if they too move to the shadow economy. Hence, the responsiveness of the skilled workers depends on their informal productivity. If the skilled workers suffer a large productivity loss by moving to the other sector, the government can tax them more in the formal sector and provide higher transfers to the unskilled informal workers. In the opposite case, however, when the skilled can easily move between sectors while the unskilled cannot, the government cannot use the shadow economy to discourage the skilled workers from reducing formal income. In such a case, redistribution will be reduced.

The shadow economy also affects the efficiency of labor allocation by sheltering workers from tax distortions.<sup>2</sup> The labor supply of formal workers is determined jointly by their formal productivity

---

<sup>1</sup>Diamond (1998) and Saez (2001) expressed the optimal tax rates in the Mirrlees model with elasticities. The higher is the elasticity of labor supply, the lower is the optimal marginal tax rate at this level of income.

<sup>2</sup>This effect corresponds to what La Porta and Shleifer (2008) call the romantic view on the shadow economy. In this view, associated with the works of Hernando de Soto (de Soto (1990, 2000)), the informal sector protects productive firms from harmful regulation and taxes.

and a marginal tax rate they face. In contrast, the labor supply of informal workers depends only on their informal production opportunity and is unaffected by tax distortions. When their informal productivity is not much lower than the formal one, informal workers will produce more than if they stayed in the formal sector. In this way the shadow economy improves the allocation of labor and raises efficiency.

Whether the shadow economy is harmful or beneficial from the social welfare perspective depends on its joint impact on redistribution and efficiency. The informal sector improves redistribution if the workers that pay high taxes cannot easily move to the shadow economy. It benefits efficiency if informal workers have similar productivities in formal and informal sector. As a rule of a thumb, we can say that the shadow economy raises welfare if it allows poor workers who collect transfers to earn some additional money, but does not tempt the rich taxpayers to reduce their formal income.

We derive the formula for the optimal tax with a shadow economy. The informal sector imposes an upper bound on the marginal tax rate, which depends on the distribution of formal and informal productivities. The optimal tax rate at each formal income level is given by either the usual Diamond (1998) formula or the upper bound, if the Diamond formula prescribes rates that are too high. In contrast to the standard Mirrlees (1971) model, in the model with shadow economy different types of workers are likely to be bunched at the single level of formal income. Specifically, all agents that supply shadow labor are subject to bunching. We develop the optimal bunching condition which complements the Diamond formula.<sup>3</sup>

The model is calibrated to Colombia, where 58% of workers are employed informally. We derive the joint distribution of formal and shadow productivity from a household survey. The main difficulty is that most individuals work only in one sector at a time. We infer their productivity in the other sector by estimating a factor: a linear combination of workers' and jobs' characteristics that explains most of the variability of shadow and formal productivities. The factor allows us to match similar individuals and infer their missing productivities. When we apply the actual tax schedule to the calibrated economy, the model replicates well the actual size of the informal sector.

We find that the optimal share of shadow workers in the total workforce is close to 22% under the Rawlsian planner and less than 1% under the Utilitarian planner. This means that the optimal shadow economy is much smaller than 58%, the actual share of shadow workers in Colombia. In comparison the Colombian income tax at the time, the optimal tax schedule has lower marginal rates at the bottom and higher rates elsewhere. Lower tax rates at the bottom displace less workers to the shadow economy, while higher tax rates above raise more revenue from high earners, yielding large welfare gains. The optimal tax rates are generally lower than the ones implied by the Mirrlees (1971) model without the informal sector. The application of the Mirrlees (1971) income tax would displace an excessive number of workers to the shadow economy.

---

<sup>3</sup>In the Mirrlees (1971) model without wealth effects the optimal allocation is described by the Diamond formula if and only if the resulting income schedule is non-decreasing, which is usually verified ex post. If the Diamond formula implies the income schedule that is decreasing at some type, our optimal bunching condition recovers the optimum.

**Related literature.** Tax evasion has been studied at least since Allingham and Sandmo (1972). For us, the most relevant paper from this literature is Kopczuk (2001). He shows that tax evasion can be welfare improving if and only if individuals are heterogeneous with respect to both productivity and tax evasion ability.<sup>4</sup> We explore this result by decomposing the welfare gain from tax evasion into the efficiency and redistribution components. Furthermore, Kopczuk (2001) derives the optimal linear income tax with tax evasion. We focus on the optimal non-linear income tax and provide a sharp characterization of the optimal shadow economy. The impact of income taxes on informal activity has been studied empirically as well. Frías, Kumler, and Verhoogen (2013) show that underreporting of wages decreases, once reported income is linked to pension benefits. Waseem (2013) documents that an increase of taxes of partnerships in Pakistan led to a massive shift to other business forms as well as a large spike in income underreporting.

Our model is focused on the workers' heterogeneity with respect to formal and informal productivities. A similar approach was taken by Albrecht, Navarro, and Vroman (2009), who study the impact of labor market institutions in a model with the formal and informal labor markets and a search friction. There is a complementary approach to modeling the shadow economy, which focuses on firms' rather than workers' heterogeneity. In Rauch (1991) managers with varying skills decide in which sector to open a business. He finds that less productive managers choose informal sector in order to avoid costly regulation. Meghir, Narita, and Robin (2015) consider heterogeneous firms that decide in which sector to operate and who are randomly matched with homogeneous workers. They find that policies aimed at reduction of the shadow economy increase competition for workers in the formal labor market and improve welfare. Amaral and Quintin (2006) to the best of our knowledge provide the only framework with the shadow economy where heterogeneity of both firms and workers is present. They extend the Rauch (1991) model by allowing for physical and human capital accumulation. Due to complementarity between the two types of capital, educated workers tend to stay in the more capital intensive formal sector.

The following two papers derive the optimal policy in related environments. Gomes, Lozachmeur, and Pavan (2014) study the optimal sector-specific income taxation when individuals can work in one of the two sectors of the economy. In our setting there are also two sectors, but the government can impose tax only on one of them. Moreover, we allow agents to work in the two sectors simultaneously. Alvarez-Parra and Sánchez (2009) study the optimal unemployment insurance with the moral hazard in search effort and an informal labor market. It is another environment with information frictions in which the informal employment is utilized in the optimal allocation.

**Structure of the paper.** In the next section we use a simple model of two types to show how the shadow economy can emerge in the optimum and what are the welfare consequences. In Section 3 we derive the optimal tax schedule with a large number of types and general social preferences. In Section 4 we introduce our methodology of extracting shadow productivities from the micro data

---

<sup>4</sup>Kopczuk (2001) describes his framework as a model of tax avoidance. In our view his results are applicable also in studying tax evasion, which is the focus of our paper.

and apply it to Colombia. We derive the optimal Colombian tax schedule in Section 5. The last section concludes.

## 2 Simple model

Imagine an economy inhabited by people that share preferences but differ in productivity. There are two types of individuals, indexed by letters  $L$  and  $H$ , with strictly positive population shares  $\mu_L$  and  $\mu_H$ . They all care about consumption  $c$  and labor supply  $n$  according to the utility function

$$U(c, n) = c - v(n). \quad (1)$$

We assume that  $v$  is increasing, strictly convex, twice differentiable and satisfies  $v'(0) = 0$ . The inverse function of  $v'$  is denoted by  $g$ .

There are two labor markets and, correspondingly, each agent is equipped with two linear production technologies. An agent of type  $i \in \{L, H\}$  produces with productivity  $w_i^f$  in a formal labor market, and with productivity  $w_i^s$  in an informal labor market. Type  $H$  is more productive in the formal market than type  $L$ :  $w_H^f > w_L^f$ . Moreover, in this section we assume that each type's informal productivity is lower than formal productivity:  $\forall_i w_i^f > w_i^s$ . We relax this assumption when we consider the full model.

Any agent may work formally, informally, or in both markets simultaneously. An agent of type  $i$  works  $n_i$  hours in total, which is the sum of  $n_i^f$  hours at the formal job and  $n_i^s$  hours in the shadow economy. The formal and the informal income, denoted by  $y_i^f$  and  $y_i^s$  respectively, is a product of the relevant productivity and the relevant labor supply. The allocation of resources may involve transfers across types, so one's consumption may be different than the sum of formal and informal income. In order to capture these flows of resources, we introduce a tax  $T_i$ , equal to the gap between total income and consumption

$$T_i \equiv y_i^f + y_i^s - c_i. \quad (2)$$

A negative tax is called a transfer, and we are going to use these terms interchangeably.

The social planner follows John Rawls' theory of justice and wants to improve the well-being of the least well-off agents,<sup>5</sup> but is limited by imperfect knowledge. The planner knows the structure and parameters of the economy, but, as in the standard Mirrlees model, does not observe the type of any individual. In addition, shadow income and labor are unobserved by the planner as well. The only variables at the individual level the planner sees and can directly verify are the formal income  $y_i^f$  and the tax  $T_i$ . We can think about  $y_i^f$  and  $y_i^f - T_i$  as a pre-tax and an after-tax reported income. Although shadow labor cannot be controlled directly, it is influenced by the choice of formal labor. Formal labor affects the marginal disutility from labor and hence changes the agent's

---

<sup>5</sup>We pick this particular point of the Pareto frontier because it allows us to show the interesting features of the model with relatively easy derivations. At the end of this section we discuss how other constrained efficient allocations look like.

optimal choice of shadow hours. Two types of labor are related according to the following function, implied by the agent's first order condition

$$n_i^s(n^f) = \max \{g(w_i^s) - n^f, 0\}. \quad (3)$$

When the agent works a sufficient number of hours in the formal sector, the marginal disutility from labor is too high to work additionally in the shadows. However, if the formal hours fall short of  $g(w_i^s)$ , the resulting gap is filled with shadow labor.

The planner maximizes the Rawlsian social welfare function, given by a utility level of the worst-off agent

$$\max_{\{(n_i^f, T_i) \in \mathbb{R}_+ \times \mathbb{R}\}_{i \in \{L, H\}}} \min \{U(c_L, n_L), U(c_H, n_H)\}, \quad (4)$$

subject to the relation between formal and shadow labor

$$n_i^s(n^f) = \max \{g(w_i^s) - n^f, 0\}, \quad (5)$$

the accounting equations

$$\forall_{i \in \{L, H\}} c_i = w_i^f n_i^f + w_i^s n_i^s(n_i^f) - T_i, \quad (6)$$

$$\forall_{i \in \{L, H\}} n_i = n_i^f + n_i^s(n_i^f), \quad (7)$$

a resource constraint

$$\sum_{i \in \{L, H\}} \mu_i T_i \geq 0, \quad (8)$$

and incentive-compatibility constraints

$$\forall_{i \in \{L, H\}} U(c_i, n_i) \geq U\left(w_{-i}^f n_{-i}^f + w_i^s n_i^s\left(\frac{w_{-i}^f}{w_i^f} n_{-i}^f\right) - T_{-i}, \frac{w_{-i}^f}{w_i^f} n_{-i}^f + n_i^s\left(\frac{w_{-i}^f}{w_i^f} n_{-i}^f\right)\right). \quad (9)$$

We denote the generic incentive constraint by  $IC_{i,-i}$ . It means that an agent  $i$  cannot be better off by earning the formal income of the other type and simultaneously adjusting informal labor.

## 2.1 First-best

What if the planner is omniscient and directly observes all variables? The planner knows types and can choose the shadow labor supply directly. The optimal allocation is a solution to the welfare maximization problem (4) where planner chooses both formal and shadow labor and a tax of each type subject only to the accounting equations (6) and (7) and the resource constraint (8). All types are more productive in the formal sector than in the shadow economy, so no agent will work informally. Each agent will supply the formal labor efficiently, equalizing the marginal social cost



and benefit of working. Moreover, the planner redistributes income from  $H$  to  $L$  in order to achieve the equality of well-being.

**Proposition 1.** *In the first-best both types work only formally and supply an efficient amount of labor:  $\forall_i v'(n_i) = w_i^f$ . Utility levels of the two types are equal:  $U(c_L, n_L) = U(c_H, n_H)$ .*

We can slightly restrict the amount of information available to the planner without affecting the optimal allocation. Suppose that the planner still observes the formal productivity, but shadow labor and income are hidden. The optimal allocation is a solution to (4) subject to the relation between shadow and formal labor (5), the accounting equations (6) and (7) and the resource constraint (8).

**Proposition 2.** *If the planner knows types, but does not observe shadow labor and income, the planner can achieve the first-best.*

When the types are known, the planner can use the lump-sum taxation and implement the first-best. Without additional frictions, the hidden shadow economy does not constrain the social planner.

## 2.2 Second-best

Let's consider the problem in which neither type nor informal activity is observed. The planner solves (4) subject to all the constraints (5) - (9). We call the solution to this problem the second-best or simply the optimum.

**Proposition 3.** *The optimum is not the first-best.  $IC_{H,L}$  is binding, while  $IC_{L,H}$  is slack.*

In the first-best, both types work only on the formal market and their utilities are equal. If  $H$  could mimic the other type, higher formal productivity would allow  $H$  to increase utility. Hence, the first-best does not satisfy  $IC_{H,L}$  and this constraint limits the welfare at the optimum. On the other hand,  $IC_{L,H}$  never binds at the optimum. It would require the redistribution of resources from type  $L$  to  $H$ , which is clearly suboptimal.

### 2.2.1 Optimal shadow economy

The standard Mirrlees model typically involves labor distortions, since they can relax the binding incentive constraints. If type  $i$  is tempted to pretend to be of the type  $-i$ , distorting number of hours of  $-i$  will discourage the deviation. Agents differ in labor productivity, so if  $i$  is more (less) productive than the other type, decreasing (increasing) number of hours worked by  $-i$  will make the deviation less attractive. Proposition 3 tells us that no agent wants to mimic type  $H$ , hence the planner has no reason to distort the labor choice of these agents. Moreover, according to (5) shadow labor is supplied only if formal labor is sufficiently distorted. Hence, the classic result of no distortions at the top implies here that  $H$  will work only formally.

**Corollary 1.** *Type H faces no distortions and never works in the shadow economy.*

On the other hand, the planner can improve social welfare by distorting the formal labor supply of type  $L$ . Stronger distortions relax the binding incentive constraint and allow the planner to redistribute more. If distortions are strong enough, type  $L$  will end up supplying shadow labor. Optimality of doing so depends on whether and by how much increasing shadow labor of type  $L$  relaxes the binding incentive constraint. As Proposition 4 demonstrates, a comparative advantage of type  $L$  in shadow labor plays a crucial role. In the proof we use the optimality condition derived in the Appendix 2 (see Lemma A.1). In order to make sure that this condition is well behaved, we require that  $v''$  is nondecreasing.<sup>6</sup>

**Proposition 4.** *Suppose that  $v''$  is nondecreasing. Type  $L$  may optimally work in the shadow economy only if*

$$\left( \frac{w_L^s}{w_L^f} - \frac{w_H^s}{w_H^f} \right) \mu_H \geq \frac{w_L^f - w_L^s}{w_L^f} \mu_L. \quad (10)$$

*Condition (10) is also a sufficient condition for type  $L$  to optimally work in the shadow economy if  $\frac{w_H^f}{w_L^f} g(w_H^s) \geq g(w_L^s)$ . Otherwise, the sufficient (but not necessary) condition is*

$$\left( \frac{w_L^s}{w_L^f} - \frac{v' \left( \frac{w_H^f}{w_H^f} g(w_L^s) \right)}{w_H^f} \right) \mu_H \geq \frac{w_L^f - w_L^s}{w_L^f} \mu_L. \quad (11)$$

Inequality (10) provides a necessary condition for the optimal shadow economy by comparing the marginal benefit and cost of increasing shadow labor of type  $L$ . The left hand side is the comparative advantage of type  $L$  over type  $H$  in the shadow labor, multiplied by the share of type  $H$ . This advantage has to be positive for type  $L$  to optimally work in the shadow economy. Otherwise, increasing shadow labor of this type does not relax the binding incentive constraint. Since the shadow economy does not facilitate screening of types, there are no benefits from the productivity-inferior shadow sector. The welfare gains from the relaxed incentive constraint are proportional to the share of type  $h$ , as the planner obtains more resources for redistribution by imposing a higher tax on this type. On the right hand side, the cost of increasing shadow labor is given by the productivity loss from using the inferior shadow production, multiplied by the share of types that supply shadow labor.

Condition (10) is also a sufficient condition for type  $L$  to work in the shadow economy if the shadow productivity of type  $H$  is not much lower than the shadow productivity of type  $L$ . If that is not the case, the optimality condition derived in Lemma A.1 is not sufficient and we have to impose a stronger sufficiency condition (11).

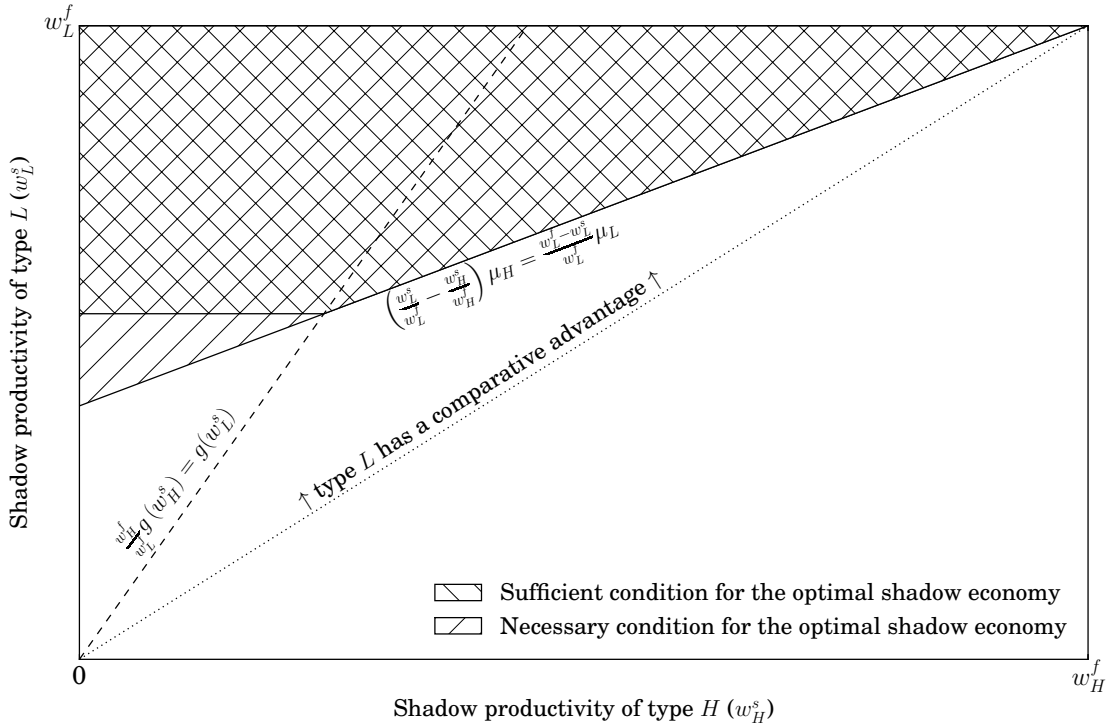
Figure 1 illustrates the proposition on the diagram of the parameter space  $(w_H^s, w_L^s)$ . Along the diagonal no type has the comparative advantage, since ratios of shadow and formal productivity of

---

<sup>6</sup>In the canonical case of isoelastic utility, it means that the elasticity of the labor supply is not greater than 1.

the two types are equal. The optimal shadow economy requires that type  $L$  has the comparative advantage in shadow labor, so the interesting action happens above the diagonal. The shadow economy is never optimal for pairs of shadow productivities which violate inequality (10). Depending on whether  $\frac{w_H^f}{w_L^f} g(w_H^s)$  is greater than  $g(w_L^s)$ , the inequality (10) is also a sufficient condition for the optimal shadow economy, or we use (11) instead. Note that the lower frontier of the necessity region crosses the vertical axis at the value  $\mu_L w_L^f$ . As the proportion of type  $L$  decreases toward zero, the region where shadow economy is optimal increases, in the limit encompassing all the points where type  $L$  has the comparative advantage over  $H$  in shadow labor.

Figure 1: The optimal shadow economy



We know when type  $L$  optimally works in the shadow economy. Proposition 5 tells us, how much shadow labor should type  $L$  supply in this case.

**Proposition 5.** *Suppose that type  $L$  optimally works in the shadow economy. Type  $L$  works only in the shadow economy if  $w_H^s \geq w_L^s$ . Type  $L$  works in both sectors simultaneously if  $w_H^s < w_L^s$ .*

When type  $L$  is more productive in the shadows than  $H$  and works only in the shadow economy, then by  $IC_{H,L}$  the utility of type  $L$  will be greater than the utility of  $H$ . Since the planner is Rawlsian, the utility levels of both types will be equalized by making type  $L$  work partly in the formal economy. On the other hand, when type  $H$  is more productive informally,  $IC_{H,L}$  means that the utility of type  $L$  will be always lower. Then if the shadow economy benefits type  $L$ , the planner will use it as much as possible.

### 2.2.2 Shadow economy and welfare

In order to examine the welfare implications of the shadow economy, we compare social welfare of the two allocations. The first one, noted with a superscript  $M$ , is the optimum of the standard Mirrlees model. We can think about the standard Mirrlees model as a special case of our model, in which both  $w_L^s$  and  $w_H^s$  are equal 0. The second allocation, noted with a superscript  $SE$ , involves type  $L$  working only in the shadow economy and the planner transferring resources from type  $H$  to type  $L$  up to the point when the incentive constraint  $IC_{H,L}$  binds. The allocation  $SE$  is not necessarily the optimum of the shadow economy model. We use it, nevertheless, to illuminate the channels through which the shadow economy influences social welfare. We measure social welfare with the utility of type  $L$ . The welfare difference between the two allocations can be decomposed in the following way

$$\underbrace{U(c_L^{SE}, n_L^{SE}) - U(c_L^M, n_L^M)}_{\text{total welfare gain}} = \underbrace{U(w_L^s n_L^{SE}, n_L^{SE}) - U(w_L^f n_L^M, n_L^M)}_{\text{efficiency gain}} + \underbrace{T_L^M - T_L^{SE}}_{\text{redistribution gain}}. \quad (12)$$

The efficiency gain measures the difference in distortions imposed on type  $L$ , while the redistribution gain describes the change in the level of transfer type  $L$  receives. Thanks to the quasilinear preferences, we can decompose these two effects additively.

**Efficiency gain.** The distortion imposed on type  $L$  in the shadow economy arise from the productivity loss  $w_L^f - w_L^s$ . By varying  $w_L^s$ , this distortion can be made arbitrarily small. On the other hand, the distortion of the standard Mirrlees model is implied by the marginal tax rate on formal income. Given redistributive social preferences, it is always optimal to impose a positive tax rate on type  $l$ . The efficiency gain, which captures the difference in distortions between two regimes, is strictly increasing in  $w_L^s$ . Intuitively, the positive efficiency gain means that the shadow economy raises social welfare by *sheltering the workers from tax distortions*.

**Redistribution gain.** The shadow economy improves redistribution if the planner is able to give higher transfer to type  $L$  (or equivalently raise higher tax from type  $H$ ). The difference in transfers can be expressed as

$$T_L^M - T_L^{SE} = \mu_H \left( U \left( w_L^f n_L^M, \frac{w_L^f}{w_H^f} n_L^M \right) - U(w_H^s n_H^{SE}, n_H^{SE}) \right). \quad (13)$$

What determines the magnitude of redistribution is the possibility of production of type  $H$  after misreporting. In the standard Mirrlees model deviating type  $H$  uses formal productivity and can produce only as much output as type  $l$ . In the allocation where type  $L$  works only informally, type  $H$  cannot supply any formal labor, but is unconstrained in supplying informal labor. Hence, the redistribution gain is strictly decreasing in  $w_H^s$ . Intuitively, a positive redistribution gain means that the shadow economy is used as a *screening device*, helping the planner to tell the types apart.

Proposition 6 uses the decomposition into the efficiency and redistribution gains in order to derive threshold values for shadow productivity of each type. Depending on which side of the thresholds the productivities are, the existence of the shadow economy improves or deteriorates social welfare in comparison to the standard Mirrlees model.

**Proposition 6.** *Define an increasing function  $H(w^s) = U(w^s g(w^s), g(w^s))$  and the following threshold values*

$$\bar{w}_L^s = H^{-1} \left( U \left( w_L^f n_L^M, n_L^M \right) \right) \in (0, w_L^f), \quad \bar{w}_H^s = H^{-1} \left( U \left( w_L^f n_L^M, \frac{w_f^f}{w_H^f} n_L^M \right) \right) \in (0, w_H^f). \quad (14)$$

*If  $w_L^s \geq \bar{w}_L^s$  and  $w_H^s \leq \bar{w}_H^s$ , where at least one of these inequalities is strict, the existence of the shadow economy improves welfare in comparison to the standard Mirrlees model.*

*If  $w_L^s \leq \bar{w}_L^s$  and  $w_H^s \geq \bar{w}_H^s$ , where at least one of these inequalities is strict, the existence of the shadow economy deteriorates welfare in comparison to the standard Mirrlees model.*

The proposition is illustrated on the Figure 2. When the shadow productivity of type  $L$  is above  $\bar{w}_L^s$ , the efficiency gain is positive. When the shadow productivity of type  $H$  is above  $\bar{w}_H^s$ , the redistribution gain is negative. Obviously, when both gains are positive (negative), the shadow economy benefits (hurts) welfare. However, the shadow economy does not have to strengthen both redistribution and efficiency simultaneously to be welfare improving. Particularly interesting is the region where the redistribution gain is negative, but the efficiency gain is sufficiently high such that the welfare is higher with the shadow economy. In this case the optimum of the shadow economy model Pareto dominates the optimum of the Mirrlees model. Type  $L$  gains, since the welfare is higher with the shadow economy. Type  $H$  benefits as well, as the negative redistribution gain implies a lower tax of this type.

### 2.2.3 General social preferences

In this short section we will derive some properties of the whole Pareto frontier of the two-types model. We consider the planner that maximizes the general utilitarian social welfare function

$$\lambda_L \mu_L U(c_L, n_L) + \lambda_H \mu_H U(c_H, n_H), \quad (15)$$

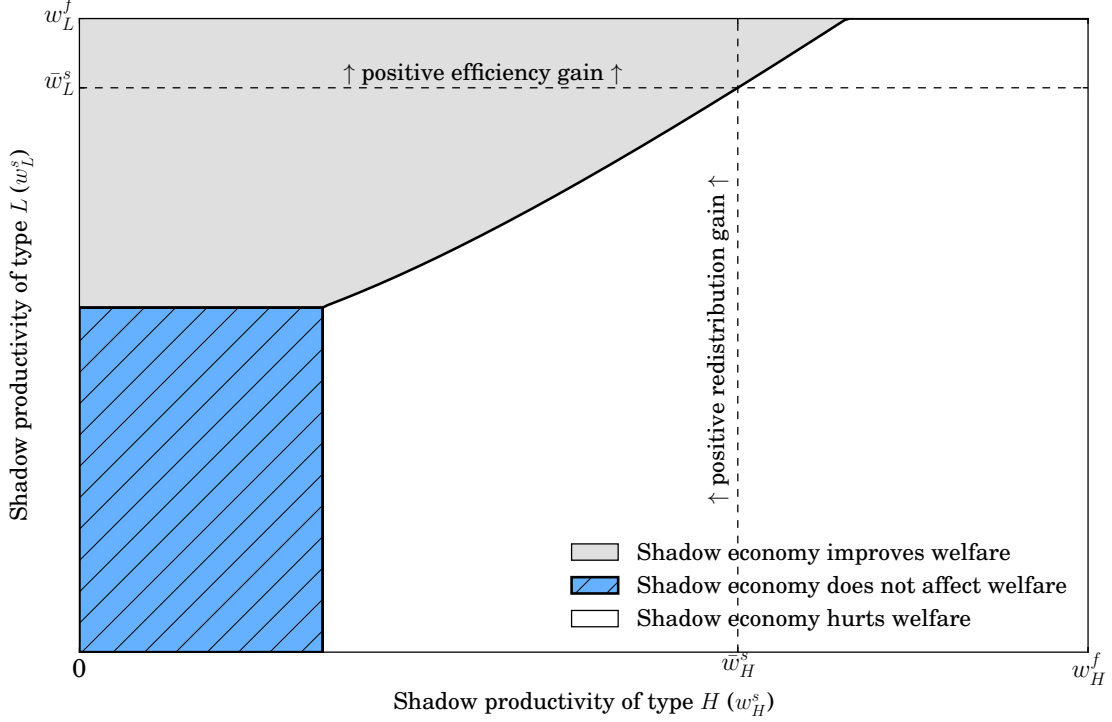
where the two Pareto weights are non-negative and sum up to 1. The maximization is subject to the constraints (5) - (9).

From the Rawlsian case we know that the comparative advantage of type  $L$  in shadow labor is necessary for this type to work in the shadows. Proposition 7 generalizes this observation.

**Proposition 7.** *Type  $i \in \{L, H\}$  may optimally work in the shadow economy only if  $\frac{w_i^s}{w_i^f} > \frac{w_{-i}^s}{w_{-i}^f}$  and  $\lambda_i > \lambda_{-i}$ .*

In order to optimally work in the shadow economy, any type  $i \in \{L, H\}$  has to satisfy two requirements. First, type  $i$  needs to have the comparative advantage in the shadow labor over the

Figure 2: Shadow economy and welfare



other type. Otherwise, shifting labor from formal to shadow sector does not relax the incentive constraints. Second, the planner has to be willing to redistribute resources to type  $i$  - the Pareto weight of this type has to be greater than the weight of the other type. The shadow economy can be beneficial only when it relaxes the binding incentive constraints, and the incentive constraint  $IC_{-i,i}$  binds if  $\lambda_i > \lambda_{-i}$ . Intuitively, if the planner prefers to tax rather than support some agents, it is suboptimal to let them evade taxation.

When will type  $i$  optimally work in the shadow economy? Let's compare the welfare of two allocations. In the first allocation (denoted by superscript  $SE$ ) type  $i$  works exclusively in the shadow economy. It provides the lower bound on welfare when type  $i$  is employed informally. The second allocation (denoted by  $M$ ) is the optimum of the standard Mirrlees model, or equivalently the optimum of the shadow economy model where  $w_i^s = w_{-i}^s = 0$ . It is the upper bound on welfare when type  $i$  is employed only in the formal sector. We can decompose the welfare difference between these two allocations in the familiar way

$$\underbrace{W^{SE} - W^M}_{\text{total welfare gain}} = \underbrace{\mu_i \lambda_i \left( U(w_i^s n_i^{SE}, n_i^{SE}) - U(w_i^f n_i^M, n_i^M) \right)}_{\text{efficiency gain}} + \underbrace{\mu_i (\lambda_i - \lambda_{-i}) (T_i^M - T_i^{SE})}_{\text{redistribution gain}}. \quad (16)$$

The welfare difference can be decomposed into the difference in effective distortions imposed on

type  $i$  and the difference in transfers received by this type. The only essential change in comparison to the simpler Rawlsian case given by (12) comes from the Pareto weights. The more the planner cares about type  $-i$ , the less valuable are gains in redistribution in comparison to the gains in efficiency.

**Proposition 8.** *Suppose that  $\lambda_i > \lambda_{-i}$  for some  $i \in \{L, H\}$ . Define the following thresholds*

$$\bar{w}_i^s = H^{-1} \left( U \left( w_i^f n_i^M, n_i^M \right) \right) \in (0, w_i^f), \quad \bar{w}_{-i}^s = H^{-1} \left( U \left( w_i^f n_i^M, \frac{w_i^f}{w_{-i}^f} n_i^M \right) \right) \in (0, w_{-i}^f). \quad (17)$$

*If  $w_i^s \geq \bar{w}_i^s$  and  $w_{-i}^s \leq \bar{w}_{-i}^s$ , where at least one of these inequalities is strict, then type  $i$  optimally works in the shadow economy and the optimum welfare is strictly higher than in the standard Mirrlees model.*

Proposition 8 generalizes the thresholds from Proposition 6. Interestingly, when the planner cares more about the more productive formally type  $H$ , these agents may end up working in the shadow economy. It may be surprising, since in the standard Mirrlees model the formal labor supply of this type is optimally either undistorted, or distorted upwards, while supplying shadow labor requires a downwards distortion. Nevertheless, if shadow economy magnifies productivity differences between types, it may be in the best interest of type  $H$  to supply only informal labor and enjoy higher transfer financed by the other type. The shadow economy in such allocation works as a tax haven, accessible only to the privileged.

### 3 Full model

In this section we describe the optimal tax schedule in the economy with a large number of types. Below we introduce a general taxation problem. Then we examine the requirements of incentive compatibility, which will involve the standard monotonicity condition. We proceed to characterize the optimal income tax. First we derive optimality conditions (which we call the *interior optimality conditions*) under the assumption that the monotonicity condition holds. It is a common practice in the literature on Mirrleesian taxation to stop here and verify the monotonicity numerically ex post. It is justified, since in the standard Mirrlees model the violation of the monotonicity requires rather unusual assumptions. On the other hand, the shadow economy provides an environment where the monotonicity condition is much more likely to be violated. We discuss in detail why it is the case and carry on to the optimality conditions when the monotonicity constraint is binding. The optimal allocation in this case involves bunching, i.e. some types are pooled together at the kinks of the tax schedule. We derive the optimal bunching condition with an intuitive variational method.<sup>7</sup> In the last subsection we summarize the main results from the full model.

<sup>7</sup>Ebert (1992) relies on the optimal control theory to derive the optimal tax when the monotonicity condition is binding. We use the more transparent variational method and develop the optimal bunching condition in the spirit of the Diamond (1998) tax formula.

### 3.1 The planner's problem

Workers are distributed on the type interval  $[0, 1]$  according to a density  $\mu_i$  and a cumulative density  $M_i$ . The density  $\mu_i$  is atomless. We assume that formal and informal productivities ( $w_i^f$  and  $w_i^s$ ) are differentiable with respect to type and denote these derivatives by  $\dot{w}_i^f$  and  $\dot{w}_i^s$ . It will be useful to denote the growth rates of productivities by  $\rho_i^x = \frac{\dot{w}_i^x}{w_i^x}$ ,  $x \in \{f, s\}$ . Types are sorted such that the formal productivity is increasing:  $\dot{w}_i^f > 0$ . We will use the dot notation to write derivatives with respect to type of other variables as well. For instance,  $\dot{y}_i^f$  stands for the derivative of formal income with respect to type, evaluated at some type  $i$ .

We focus on preferences without wealth effects. Agents' utility function is  $U(c, n) = c - v(n)$ , where  $v$  is increasing, strictly convex and twice differentiable function. We denote the inverse function of the marginal disutility from labor  $v'$  by  $g$  and the elasticity of labor supply of type  $i$  by  $\zeta_i$ .<sup>8</sup> Let  $V_i(y^f, T)$  be the indirect utility function of an agent of type  $i$  whose reported formal income is  $y^f$  and who pays a tax  $T$ :

$$V_i(y^f, T) \equiv \max_{n^s \geq 0} y^f + w_i^s n^s - T - v\left(\frac{y^f}{w_i^f} + n^s\right). \quad (18)$$

In addition to earning the formal income, the agent is optimally choosing the amount of informal labor. Due to concavity of the problem, the choice of  $n^s$  is pinned down by the familiar first order condition, modified to allow for the corner solution

$$\min \left\{ v' \left( \frac{y^f}{w_i^f} + n^s \right) - w_i^s, n^s \right\} = 0. \quad (19)$$

Whenever the formal income  $y^f$  is sufficiently high, no shadow labor is supplied. Conversely, sufficiently low formal income leads to informal employment.

The planner chooses a formal income schedule  $y^f$  and a tax schedule  $T$  in order to maximize a general social welfare function

$$\max_{(y_i^f, T_i)_{i \in [0, 1]}} \int_0^1 \lambda_i G(V_i(y_i^f, T_i)) d\mu_i, \quad (20)$$

where  $G$  is an increasing and differentiable function and the Pareto weights  $\lambda \in [0, 1] \rightarrow \mathbb{R}_+$  integrate to 1.<sup>9</sup> The budget constraint is the following

$$\int_0^1 T_i d\mu_i \geq E, \quad (21)$$

<sup>8</sup>Since we abstract from wealth effects, the compensated and uncompensated elasticities coincide. Note that the elasticity is in general an endogenous object, as it depends on labor supply:  $\zeta_i = \frac{v'(n_i)}{n_i v''(n_i)}$ .

<sup>9</sup>It's easy to relax the assumption of a finite Pareto weight on each type and we are going to do it in the quantitative section, where we consider, among others, the Rawlsian planner.



where the net tax revenue needs to cover some fixed expenditures  $E$ . Moreover, the tax schedule has to satisfy incentive compatibility

$$\forall_{i,j \in [0,1]} V_i(y_i^f, T_i) \geq V_i(y_j^f, T_j), \quad (22)$$

which means that no agent can gain by mimicking any other type. The allocation which solves (20) subject to (21) and (22) is called the second-best or the optimum.

We will describe the optimum by specifying the marginal tax rate of each type. The marginal tax rate is given by the ratio of slopes of the total tax schedule and the formal income schedule

$$t_i = \frac{\dot{T}_i}{\dot{y}_i^f}. \quad (23)$$

Intuitively, it describes the fraction of a marginal formal income increase that is claimed by the planner.

### 3.2 Incentive-compatibility

The single crossing property allows the planner in the standard Mirrlees model to focus only on local incentive compatibility constraints. Intuitively, the single-crossing means that, given a constant tax rate, a higher type is willing to earn more than a lower type. The single-crossing in our model means that, holding the tax rate constant, the higher type is willing to earn *formally* more than the lower type.

**Assumption 1.** *A comparative advantage in shadow labor is decreasing with type:  $\frac{d}{di} \left( \frac{w_i^s}{w_i^f} \right) < 0$ .*

**Lemma 1.** *Under Assumption 1, the indirect utility function  $V$  has the single crossing property.*

The single-crossing holds when the agents with lower formal productivity have a comparative advantage in working in the informal sector. The single-crossing allows us to replace the general incentive compatibility condition (22) with two simpler requirements.

**Proposition 9.** *Under Assumption 1, the allocation  $(y_i^f, T_i)_{i \in [0,1]}$  is incentive-compatible if and only if the two conditions are satisfied:*

1.  $y_i^f$  is non-decreasing in type.
2. If  $\dot{y}_i^f$  exists, then the local incentive-compatibility condition holds:  $\frac{d}{dj} V_i(y_j^f, T_j) \Big|_{j=i} = 0$ .

The utility schedule  $V_i(y_i^f, T_i)$  of an incentive compatible allocation is continuous everywhere, differentiable almost everywhere and for any  $i < 1$  can be expressed as

$$V_i(y_i^f, T_i) = V_0(y_0^f, T_0) + \int_0^i \dot{V}_j(y_j^f, T_j) dj, \quad (24)$$

where

$$\dot{V}_j(y_j^f, T_j) \equiv (\rho_j^f n_j^f + \rho_j^s n_j^s) v'(n_j). \quad (25)$$

The single crossing implies that for any tax schedule the level of formal income chosen by a worker is weakly increasing in the worker's type. Hence, assigning a lower income to a higher type would violate incentive compatibility. It is enough to focus just on local deviations: no agent should be able to improve utility by marginally changing the formal earnings. This local incentive-compatibility constraint is equivalent to the familiar condition for the optimal choice of the formal income given the marginal tax rate  $t_i$ , allowing for the corner solution

$$\min \left\{ v' \left( \frac{y_i^f}{w_i^f} + n_i^s \right) - (1 - t_i) w_i^f, y_i^f \right\} = 0. \quad (26)$$

Note that the formal income may be, and sometimes will be, discontinuous in type. Nevertheless, the indirect utility function preserves some smoothness and can be expressed as an integral of its marginal increments.

Let's call  $\dot{V}_i(y_i^f, T_i)$  the *marginal information rent* of type  $i$ . It describes how the utility level changes with type. The higher the average rate of productivity growth, weighted by the labor inputs in two sectors, the faster utility increases with type. We will use perturbations in the marginal information rent to derive the optimal tax schedule.

In what follows we will economize on notation of the utility schedule and its slope by supressing the arguments:  $V_i \equiv V_i(y_i^f, T_i)$  and  $\dot{V}_i \equiv \dot{V}_i(y_i^f, T_i)$ .

### 3.3 Optimality conditions

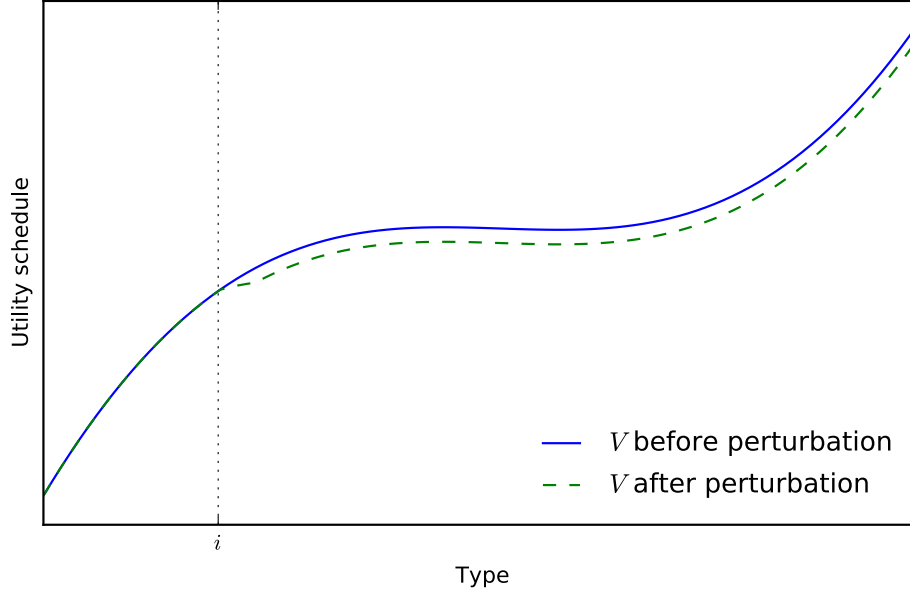
First, we solve for the optimum under assumption that the resulting formal income schedule is non-decreasing. Second, we examine when this assumption is justified and show that the existence of the shadow economy make it's violation more likely. Finally, we derive the optimality conditions in the general case.

#### 3.3.1 Interior optimality conditions

We obtain the interior optimality conditions by making sure that the social welfare cannot be improved by perturbing the marginal information rent of any type.<sup>10</sup> A marginal information rent is a slope of the utility schedule at some type  $i$ . It can be reduced by increasing tax distortions of this type, which is costly for the budget. On the other hand, by (24) such perturbation shifts downwards the entire utility schedule above type  $i$  (see Figure 3). This shift is a uniform increase of a non-distortionary tax of all types above  $i$ . The interior optimality conditions balance the

<sup>10</sup>To the best of our knowledge, Brendon (2013) was the first to use this approach in the Mirrlees model. He also inspired us to express the optimality conditions with endogenous cost terms, although our notation differs from his.

Figure 3: Decreasing the marginal information rent of type  $i$



cost of distortions with gains from efficient taxation for each type. Below we present terms that capture the marginal costs and benefits of such perturbations. We derive them in detail in the proof of Theorem 1. The shadow economy enters the picture by affecting the cost of increasing tax distortions.

The benefit of shifting the utility schedule of type  $j$  without affecting its slope is given by the standard expression

$$N_j \equiv (1 - \omega_j) \mu_j, \text{ where } \omega_j = \frac{\lambda_j}{\eta} G'(V_j). \quad (27)$$

A marginal increase of non-distortionary taxation of type  $j$  leads to one-to-one increase of tax revenue. On the other hand, it reduces the social welfare, since the utility of type  $j$  falls. Following Piketty and Saez (2013) we call this welfare impact the *marginal welfare weight* and denote it by  $\omega_j$ . Note that welfare impact is normalized by the Lagrange multiplier of the resource constraint  $\eta$ . It allows us to express changes in welfare in the unit of resources. We multiply the whole expression by the density of type  $j$  in order to include all agents of this type. We assumed that there are no wealth effects, so the non-distortionary tax does not affect the labor choice of agents. Consequently, the term  $N_j$  does not depend on whether type  $j$  works informally.

The cost of decreasing some agent's marginal information rent depends on the involvement of this

agent in the shadow activity. Types can be grouped into three sets:

$$\begin{aligned} \text{formal workers:} \quad & \mathcal{F} \equiv \left\{ i \in [0, 1] : v' \left( n_i^f \right) > w_i^s \right\}, \\ \text{marginal workers:} \quad & \mathcal{M} \equiv \left\{ i \in [0, 1] : v' \left( n_i^f \right) = w_i^s \right\}, \\ \text{shadow workers:} \quad & \mathcal{S} \equiv \left\{ i \in [0, 1] : v' \left( n_i^f \right) < w_i^s \right\}. \end{aligned}$$

The formal workers supply only formal labor: their marginal disutility from working is strictly greater than their shadow productivity. The marginal workers also supply only formal labor, but their marginal disutility from work is exactly equal to their shadow productivity. A small reduction of formal labor supply of these agents would make them work in the informal sector. Finally, the shadow workers are employed informally, although they can also supply some formal labor.

The formal workers act exactly like agents in the standard Mirrlees model. By increasing distortions, the planner is reducing their total labor supply. The cost of increasing distortions is given by

$$D_i^f \equiv \frac{t_i}{1-t_i} \left( \rho_i^f \left( 1 + \frac{1}{\zeta_i} \right) \right)^{-1} \mu_i. \quad (28)$$

The cost depends positively on the marginal tax rate. The marginal tax rate tell us how strongly a reduction of the formal income influences the tax revenue. Moreover, the cost increases with the elasticity of labor supply  $\zeta_i$  and is proportional to the density of the distorted type.  $D_i^f$  is endogenous, as it depends on the marginal tax rate.

The perturbation of the marginal information rent works differently for the shadow workers. They supply shadow labor in the quantity that satisfies  $v' \left( n_i^f + n_i^s \right) = w_i^s$ , which means that their total labor supply  $n_i$  is constant. By distorting the formal income, the planner simply shift their labor from the formal to the informal sector. As a result, the cost of increasing distortions does not depend on the elasticity of labor supply, but rather on the sectoral productivity differences,

$$D_i^s \equiv \frac{w_i^f - w_i^s}{w_i^s} \left( \rho_i^f - \rho_i^s \right)^{-1} \mu_i. \quad (29)$$

The first term is the relative productivity difference between formal and informal sector. Actually, it's also equal to  $\frac{t_i}{1-t_i}$ , since the marginal tax rate of these types equalizes the return to labor in both sectors:  $(1-t_i) w_i^f = w_i^s$ . Hence, as in the case of formal workers, the first term corresponds to the direct tax revenue cost of reduced formal labor supply. The second term describes how effectively the planner can manipulate the agent's marginal information rent by discouraging the formal labor. By the single-crossing assumption, this term is always positive. Again, the density  $\mu_i$  aggregates the expression to include all agents of type  $i$ . Note that  $D_i^s$  is exogenous, as it depends only on the fundamentals of the economy.

The marginal workers are walking a tightrope between their formal and shadow colleagues. If the planner marginally reduces their income, they become the shadow workers. If the planner lifts

distortions, they join the formal workers. The cost of changing distortions of these types depends on the direction of perturbation and is equal to either  $D_i^f$  or  $D_i^s$ .

Having all the cost and benefit terms ready, we can derive the interior optimality conditions. Recall, that by varying the distortions imposed on some type, the planner changes a non-distortionary tax of all types above. In the optimum, the planner cannot increase the social welfare by such perturbations. For the formal workers, this means that

$$\forall_{i \in \mathcal{F}} D_i^f = \int_i^1 N_j dj. \quad (30)$$

It is a standard optimality condition from the Mirrlees model, derived first in the quasilinear case by Diamond (1998). The shadow economy does not affect the marginal tax rate of formal agents directly. It may influence them only indirectly, by changing the marginal welfare weights of types above.

For the marginal workers it must be the case that increasing tax distortions is beneficial as long as they work only formally, but it is too costly when they start to supply the shadow labor.

$$\forall_{i \in \mathcal{M}} D_i^s \geq \int_i^1 N_j dj \geq D_i^f \text{ and } y_i^f = w_i^f g(w_i^s). \quad (31)$$

The marginal workers do not supply informal labor, but in their case the shadow economy constitutes a binding constraint for the planner. Absent the shadow economy, the marginal tax rates would be set at a higher level. In our model the planner is not willing to do it, because it would push the marginal workers to informal jobs, which is too costly. Formal labor supply of the marginal workers is fixed at the lowest level that leaves them no incentives to work informally.

Recall that the cost of distorting the shadow worker is fixed by the parameters of the economy. Moreover, the benefit of distorting one particular worker, given by (27), is fixed as well, since the perturbation of the marginal information rent of  $i$  has an infinitesimal effect on the utility of types above. If the planner finds it optimal to decrease the formal income of agent  $i$  so much that  $i$  starts supplying informal labor, it will be optimal to decrease the formal income all the way to zero, when  $i$  works only in the shadow economy:

$$\forall_{i \in \mathcal{S}} \int_i^1 N_j dj > D_i^s \text{ and } y_i^f = 0. \quad (32)$$

Note that according to this condition all shadow workers are bunched together at zero formal income.<sup>11</sup>

---

<sup>11</sup>Notice that we could replace the strict inequality with a weak one in (32), and conversely regarding the left inequality in (31). In words, when the cost of distorting some marginal worker is exactly equal to the benefit, then this worker could equally well be a shadow worker, with no change in the social welfare. It means that whenever the curves  $D_i^s$  and  $\int_i^1 N_j dj$  cross, the optimum is not unique, since we could vary allocation of the type at the intersection. Since such a crossing is unlikely to happen more than a few times, we do not consider this as an important issue. We sidestep it by assuming that the planner introduces distortions only when there are strictly positive gains from doing so. Consequently, our notion of uniqueness of optimum should be understood with this reservation.

The optimality conditions (30)-(32) determine the slope of the utility schedule at each type. What is left is finding the optimal level. Suppose that the planner varies the tax paid by the lowest type, while keeping all the marginal rates fixed. Optimum requires that such perturbation cannot improve welfare:

$$\int_0^1 N_j dj = 0. \quad (33)$$

**Definition.** The conditions (30)-(33) are called the *interior optimality conditions*. The allocation  $(y^f, T)$  consistent with the interior optimality conditions is called the *interior allocation*. Specifically,  $y^f$  is called the *interior formal income schedule*.

The interior conditions are necessary for the optimum as long as they don't imply a formal income schedule which is locally decreasing. They become sufficient, if they pin down a unique allocation. This happens when the cost of distortions is increasing in the amount of distortions imposed. When that is the case, the planner's problem with respect to each type becomes concave. Theorem 1 provides regularity conditions which guarantee it.

**Assumption 2.** (i) The elasticity of labor supply  $\frac{v'(n)}{nv''(n)}$  is non-increasing in  $n$ . (ii) The ratio of sectoral growth rates is bounded below  $\forall_i \frac{\rho_i^s}{\rho_i} > -\zeta_i^{-1}$ .

**Theorem 1.** Under Assumption 1, if all interior formal income schedules are non-decreasing, the interior optimality conditions are necessary for the optimum. Under Assumptions 1 and 2, there is a unique interior formal income schedule. If it is non-decreasing, the interior optimality conditions are both necessary and sufficient for the optimum.

### 3.3.2 When do the interior conditions fail?

The interior allocation is incentive-compatible and optimal if it leads to formal income that is non-decreasing in type. In the standard Mirrlees model formal income is decreasing if the marginal tax rate increases too quickly with type. However, in virtually all applications of the standard Mirrlees model this is not a problem, as the conditions under which the interior tax rate increases that fast are rather unusual.<sup>12</sup> The shadow economy gives rise to another reason for non-monotone interior formal income. In the interior allocation all shadow workers have zero formal income. Hence, if there is any worker with positive formal income with a type lower than some shadow worker, the formal income schedule will be locally decreasing. It turns out that this second reason makes the failure of the interior allocation much more likely. In Proposition 10 below we provide the sufficient conditions for the formal income to be non-decreasing. Then we discuss the two cases in which the shadow economy leads to the failure of the interior optimality conditions.

**Assumption 3.** (i) The social welfare function is such that  $G(V) = V$ ,  $\lambda_i$  is non-decreasing in type for  $i > 0$ . (ii) The ratio  $\frac{1}{\rho_i} \frac{\mu_i}{1-M_i}$  is non-decreasing in type. (iii) The elasticity of labor supply is constant:  $\forall_i \zeta_i = \zeta$ . (iv) The ratio of sectoral growth rates  $\frac{\rho_i^s}{\rho_i}$  is non-decreasing in type.

<sup>12</sup>Probably simplest way to construct an example of locally decreasing formal income schedule is to assume a bimodal productivity distribution, with very low density between the modes.

**Proposition 10.** *Under Assumptions 1, 2 and 3, the unique interior formal income schedule is non-decreasing.*

First, notice that we make sure that the interior formal income schedule is unique (Assumption 2). Simultaneously, it implies that the formal income of the marginal workers is non-decreasing. Assumptions 3(i) - 3(iii) make sure that the marginal tax rate of formal workers is non-increasing in type, which in turn implies that the formal income of these workers is non-decreasing. These conditions are familiar from the standard Mirrlees model. Assumption 3(i) is satisfied by the utilitarian or Rawlsian social welfare function, while Assumption 3(ii) is a weaker counterpart of the usual monotone hazard ratio requirement.<sup>13</sup>

Finally, we have to make sure that all shadow workers, if there are any, are at the bottom of the type space. By (32) it means that the marginal cost of distorting the shadow worker  $D_i^s$  can cross the marginal benefit  $\int_i^1 N_j dj$  at most once and from below. It is guaranteed jointly by conditions 3(i), 3(ii) and the new requirement 3(iv) which says that the ratio of sectoral productivity growth rates is non-decreasing. In addition to assuring the optimality of the interior allocation, Assumption 3 imply also that sets  $\mathcal{S}, \mathcal{M}$  and  $\mathcal{F}$ , if non-empty, can be ordered: the bottom types are the shadow workers, above them are the marginal workers, and the top types are formal.

Assumption 2 makes sure that the  $D_i^s$  curve crosses the  $\int_i^1 N_j dj$  curve at most once. Let's see how the relaxation of some of its elements make these curves cross more than once. In Example 1 we relax the assumption on the social welfare function and in Example 2 we allow the non-monotone ratio of sectoral growth rates.

**Example 1.** (i) The social welfare function is such that  $G(V) = V$ , the Pareto weights  $\lambda_i$  are continuous in type and satisfy  $\lambda_0 > 2$ . (ii) The distribution of types is uniform. (iii) The elasticity of labor supply is constant:  $\forall_i \zeta_i = \zeta$  and  $v'(0) = 0$ . (iv) The ratio of sectoral growth rates  $\frac{\rho_i^s}{\rho_i^f}$  is fixed. (v) Assumptions 1 and 2 are satisfied.

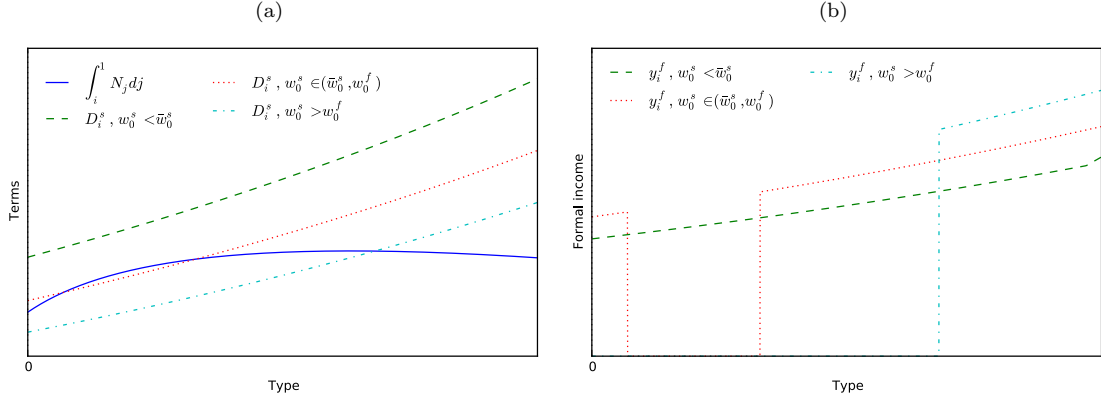
**Lemma 2.** *In Example 1 there is a threshold  $\bar{w}_0^s \in (0, w_0^f)$  such that if  $w_0^f > w_0^s > \bar{w}_0^s$  the interior formal income schedule **is not** non-decreasing.*

Example 1 violates Assumption 3 (i), which allows the  $\int_i^1 N_j dj$  term to be initially increasing in type.<sup>14</sup> Both terms  $D_i^s$  and  $\int_i^1 N_j dj$  are increasing at 0, but  $\int_i^1 N_j dj$  term increases faster. If  $w_0^f > w_0^s$ , then the distortion cost at type 0 is greater than the benefit and the bottom type works formally. If the gap between  $w_0^f$  and  $w_0^s$  is sufficiently small (smaller than  $w_0^f - \bar{w}_0^s > 0$ ),  $D_i^s$  curve will cross the benefit curve at some positive type (see Figure 4). Consequently, the agents above the intersection will work in the shadow economy. Since these agents have no formal income, the formal income schedule is locally decreasing.

<sup>13</sup>We can express the distribution of types as a function of formal productivity rather than type. Then the density is  $\bar{\mu}(w_i^f) = \frac{\mu_i}{w_i^f}$  and cumulative density is  $\bar{M}(w_i^f) = M_i$ . Hence, assumption 3(ii) means that  $\frac{w^f \bar{\mu}(w^f)}{1 - \bar{M}(w^f)}$  is non-decreasing. For instance, any Pareto distribution of formal productivity satisfies this assumption.

<sup>14</sup>The Pareto weights integrate to 1 over the type space, so they have to be lower than or equal to 1 for some types above 0. Since these weights are continuous and  $\lambda_0 > 2$ , they will be decreasing for some type above 0, violating 3(i).

Figure 4: A failure of the interior allocation due to increasing benefit of distortions  $\int_i^1 N_j dj$  (Example 1).



**Example 2.** (i) The social welfare function is Rawlsian:  $\forall_{i>0} \lambda_i = 0$ . (ii) The distribution of types is uniform. (iii) The elasticity of labor supply is constant:  $\forall_i \zeta_i = \zeta$ . (iv) The growth rate of formal productivity is fixed, while the growth rate of shadow productivity is decreasing for some types. (v) Assumptions 1 and 2 are satisfied.

Example 2 satisfies all the requirements of Proposition 10 apart from the non-decreasing sectoral growth rates ratio assumption. In panel (a) of Figure 5 we can see that the growth rate of shadow productivity decreases around the middle type and then bounces back. It is reflected in the marginal cost of distorting shadow workers  $D_i^s$  (panel (b)). We chose the parameters such that the fall is substantial, making the  $D_i^s$  curve cross the  $\int_i^1 N_j dj$  curve three times. Consequently, the formal income first increases, then decreases to 0 once the  $D_i^s$  crosses  $\int_i^1 N_j dj$  for the second time. This example shows that even minor irregularities in the distribution of productivities can make the interior allocation not implementable.

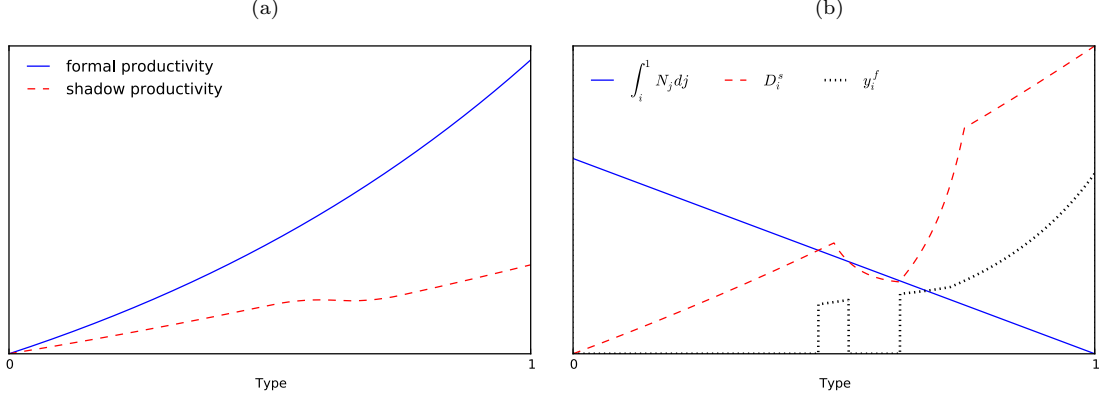
### 3.3.3 Optimal bunching

Whenever the interior formal income schedule is decreasing for some types, the interior allocation is not incentive-compatible and hence is not optimal. Ebert (1992) and Boadway, Cuff, and Marchand (2000) applied the optimal control theory to overcome this problem. In contrast to these papers, we derive the optimal bunching condition with the intuitive variational argument and express it in the spirit of the Diamond (1998) optimal tax formula. What we are going to do is essentially “ironing” the formal income schedule whenever it is locally decreasing (see Figure 6). The ironing was originally introduced by Mussa and Rosen (1978) in a solution to the monopolistic pricing problem when the monotonicity condition is binding.

Suppose that the interior formal income schedule  $\bar{y}^f$  is decreasing on some set of types, beginning with  $\bar{a}$ . Decreasing formal income is incompatible with the incentive-compatibility. We can regain



Figure 5: A failure of the interior allocation due to non-monotone ratio of productivity growth rates (Example 2).



incentive-compatibility by lifting the schedule such that it becomes overall non-decreasing and flat in the interval  $[\bar{a}, \bar{b}]$  (see Figure 6). Since types  $[\bar{a}, \bar{b}]$  have the same formal income, they are bunched and cannot be differentiated by the planner. Such bunching is implemented by a discontinuous jump of the marginal tax rate.

The flattened schedule is incentive-compatible. However, generally it is not optimal. By marginally decreasing formal income of type  $\bar{a}$  the planner relaxes the binding monotonicity constraint and can marginally decrease the formal income of all types in the interval  $(\bar{a}, \bar{b})$ . This perturbation closes the gap between the actual formal income and its interior value for the positive measure of types. On the other hand, the cost of perturbation is infinitesimal: it is a distortion of one type  $\bar{a}$ . This perturbation is clearly welfare-improving, starting from the flattened interior schedule. Below we find the optimal bunching condition by making sure that the perturbation is not beneficial at the optimal income schedule.

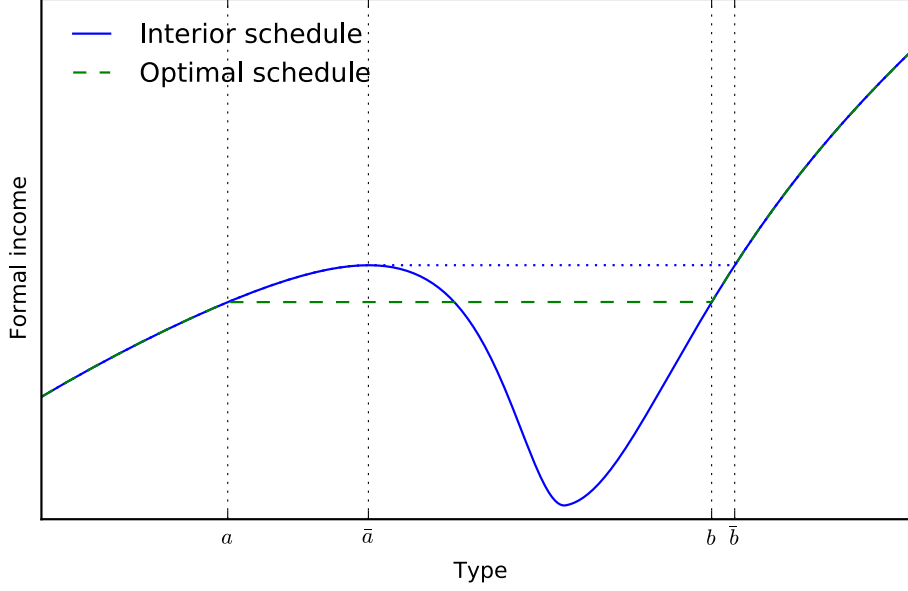
Suppose that an interval of agents  $[a, b]$  is bunched. Let's marginally decrease the formal income of agents  $[a, b]$  and adjust their total tax paid such that the utility of type  $a$  is unchanged. In this way we preserve the continuity of the utility schedule. However, since the other bunched agents have a different marginal rate of substitution between consumption and income, this perturbation will decrease their utility. We normalize the perturbation such that we obtain a unit change of the utility of the highest type in the bunch. The total cost of this perturbation is given by

$$D_{a,b} \equiv (t_a + \mathbb{E} \{ \Delta MRS_i \omega_i | b > i \geq a \}) \frac{M_b - M_a}{t_{b+} - t_{a-}}, \quad (34)$$

$$\text{where } \Delta MRS_i = \frac{v'(n_a)}{w_a^f} - \frac{v'(n_i)}{w_i^f}.$$

The expression within the brackets is an average impact of a unit perturbation of the formal income. The brackets contain two components: a fiscal and a welfare loss. The fiscal loss from reducing the

Figure 6: Ironing the formal income schedule



formal income of each bunched agent is the marginal tax rate below the kink. The welfare loss is an average marginal welfare weight in the bunch corrected by a discrepancy of the marginal rate of substitution of a given type from type  $a$ . The larger  $\Delta MRS_j$  is, the more type  $j$  suffers from the perturbation. Note that  $\Delta MRS_b$  is just equal  $t_{b+} - t_{a-}$ .<sup>15</sup> Hence, in order to normalize the perturbation to have a unit impact on utility of type  $b$ , we divide the brackets by  $t_{b+} - t_{a-}$ . We aggregate this average effect by multiplying it by the mass of bunched types.

The benefit of this perturbation comes from the reduced utility of types above  $b$  and is the same as in the interior case. The optimality requires that

$$\min \left\{ \int_b^1 N_j dj - D_{a,b}, y_a^f \right\} = 0. \quad (35)$$

Note that the optimality condition involves a corner solution when  $y_a^f = 0$ . It corresponds to the situation in which the bunched workers don't work formally at all.

The optimality condition (35) is influenced by the shadow economy again through the cost of distortion. If some worker  $i$  in the bunch  $[a, b]$  supplies shadow labor, then the difference in the marginal rate of substitution for this worker is given by  $\Delta MRS_i = \frac{v'(n_a)}{w_a^f} - \frac{w_i^s}{w_i^f}$ .

Theorem 2 combines all the optimality conditions.

<sup>15</sup>The marginal tax rate discontinuously increases at the kink. By  $t_{a-}$  we denote the tax rate below the kink and by  $t_{b+}$  the tax rate above the kink.

**Theorem 2.** *Under Assumption 1, the optimal allocation satisfies (33) and at each level of formal income one of the three mutually exclusive alternatives hold:*

- *there is no type that reports such formal income,*
- *there is a unique type whose allocation satisfies the interior optimality conditions (30)-(32),*
- *there is a bunch of types whose allocation satisfy the optimal bunching condition (35).*

Although we managed to characterize the full set of optimality conditions, the interior conditions are generally easier to use. Below we show that the interior allocation, even if not incentive-compatible, are a good predictor of which agents optimally work in the shadow economy.

**Assumption 4.** *(i)  $G$  is a concave function. (ii)  $\rho_i^f$ ,  $\rho_i^s$ ,  $\mu_i$  and  $\lambda_i$  are continuous in type.*

**Proposition 11.** *Under Assumptions 1, 2 and 4, all the types that supply shadow labor in the interior allocation remain the shadow workers in the optimum.*

### 3.4 Summary of results

**Which agents should work in the shadow economy?**

**Corollary 2.** *Suppose that  $v'(0) = 0$ . Under Assumptions 1, 2 and 4 type  $i$  optimally works in the shadow economy if*

$$\mathbb{E} \{ 1 - \omega_j | j > i \} \geq \frac{w_i^f - w_i^s}{w_i^f} \left( -\frac{d}{di} \left( \frac{w_i^s}{w_i^f} \right) \right)^{-1} \frac{\mu_i}{1 - M_i}. \quad (36)$$

*This condition is both necessary and sufficient if the interior allocation is incentive-compatible.*

The inequality (36) compares the gains from efficient taxation of all types above  $i$  with the cost of distorting type  $i$ , when this type is at the edge of joining the shadow economy. A type  $i$  is likely to optimally work in the shadow economy if the planner on average puts a low marginal welfare weights on the types above  $i$ , the relative productivity loss from moving to informal employment is low and the density of distorted types is low in comparison to the fraction of types above. Finally, the shadow employment is more likely if the comparative advantage of working in the shadow sector  $\frac{w_i^s}{w_i^f}$  is quickly decreasing with type. It means that higher types have less incentives to follow type  $i$  into the shadow economy. We assume  $v'(0) = 0$  so that we do not have to worry about some types not supplying any labor at all.

Note that with the Rawlsian planner the inequality (36) is just a continuous equivalent of the condition (10) from the simple model.

**The optimal tax rates.** Let's focus on agents that supply some formal labor and are not bunched at the kinks of the tax schedule. These types never supply informal labor. The optimal tax formula is

$$\frac{t_i}{1-t_i} = \min \left\{ \frac{w_i^f - w_i^s}{w_i^s}, \rho_i^f \left( 1 + \frac{1}{\zeta_i} \right) \frac{1 - M_i}{\mu_i} \mathbb{E}(1 - \omega_j | j > i) \right\}. \quad (37)$$

The shadow economy imposes an upper bound on the marginal tax rate. The bound (the left term in the min operator of (37)) is such that the tax rate equalizes the return from formal and informal labor - it is the highest tax rate consistent with agents working in the formal sector.

If the bound is not constraining the planner, then the tax rate should be set according to Diamond (1998) formula (the right term in the min operator of (37)). The expectations describe the average social preferences towards all types above  $i$ . In general, the less the planner cares about increasing utility of the types above  $i$ , the higher  $t_i$  will be. If the Pareto weights increase with type or  $G$  is a strictly convex function, this term may become negative, leading to negative marginal tax rates, as explained by Choné and Laroque (2010). Since the sign of the tax rate is ambiguous, below we describe how the other terms influence its absolute value. The optimal tax rate increases in absolute value when the growth rate of formal productivity with respect to type is high. If the planner is redistributive and types above  $i$  are much more productive than types below, it is optimal to set a high tax rate. The tax rate decreases with elasticity of labor supply  $\zeta_i$ , as it makes workers more responsive to the tax changes. The ratio  $\frac{1-M_i}{\mu_i}$  tells us how many agents will be taxed in a non-distortionary manner relative to the density of distorted agents. If this ratio is high, the gain from increasing tax rates relative to the cost will be high as well.

**Optimal bunching.** Bunching may arise at the bottom of the formal income distribution, resulting in de facto exclusion from the formal labor market. Bunching may also appear at a positive level of formal income, which implies a kink in a tax schedule. All workers who supply shadow labor are subject to bunching, though not necessarily at the same tax kink. Some workers supplying only formal labor can be found at the kinks as well. The formal income schedule at which the kink is located is determined by

$$\frac{t_{a-}}{t_{b+} - t_{a-}} = \frac{1 - M_b}{M_b - M_a} \mathbb{E}\{1 - \omega_j | j \geq b\} - \mathbb{E}\left\{ \frac{\Delta MRS_i}{\Delta MRS_b} w_i \middle| b > i \geq a \right\}, \quad (38)$$

where  $a$  and  $b$  are respectively the lowest and the highest type bunched at the kink. Note that both  $t_{a-}$  and  $t_{b+}$ , the tax rates below and above the kink, are set according to (37). The location of the kink is determined by the trade-off between tax and welfare losses from the bunched agents and the tax revenue gains from the efficient taxation of agents above the kink.

## 4 Measuring shadow and formal productivities

To assess the practical relevance of our theoretical results we proceed to look at the empirical counterparts of the building blocks of our theory. We focus on a developing economy with a large shadow sector: Colombia.<sup>16</sup> In this section we empirically estimate the three key objects of the model: the formal productivity ( $w_i^f$ ), the informal productivity ( $w_i^s$ ) and the distribution of types ( $\mu_i$ ). In section 5 we use our estimates to analyze how the existence of the shadow economy shapes the optimal tax scheme in Colombia.

Colombia is a case that suits itself very well to take our theory to the data, because the shadow economy is large and we can actually observe the total income of individuals, both if formal or shadow, through survey data. Household surveys reveal information about shadow income without making it usable by the authorities to levy taxes.<sup>17</sup> Furthermore, Colombian regulation makes it easy to infer shadow and formal income from questions about total income, and from the type of affiliation of the worker to the social security system.

In the model,  $w_i^f$  and  $w_i^s$  correspond to the pre-tax (real) income for one unit of labour for individual of type  $i$  in each sector, and  $\mu_i$  is the density of such type. Therefore, we have one-dimensional heterogeneity across individuals. Our empirical strategy is to replicate such one-dimensional heterogeneity by using a factor that comprises information of the worker and job characteristics, such as the education level and the task done on the job. The identification assumptions is that the pre-tax hourly wage recorded on the surveys is a noisy signal of the productivities in each sector and that the productivities themselves are a linear function of the factor we employ.

The weights that are used to construct the factor and the parameters that map productivities to wages are jointly estimated to maximize the explanatory content of the factor over wages. Indeed, the factor we obtain can explain most of the variability of wages in both sectors. Nevertheless, the factor cannot account for the income dispersion of the top earners and the gap with respect to the rest of the population. We extend our identification strategy by estimating a Pareto distribution for the wages of top earners in the formal sector.

We find that both productivity estimates are increasing in type (the factor) and that the single-crossing property is satisfied. Specifically, the wedge between the productivity levels of each sector is almost zero for the least productive agents and increases rapidly as the formal productivity increases. The main novelty of this section is that we assess the differences between the formal and the shadow economy at the worker level, controlling for the sorting of workers. Productivity as measured in La Porta and Shleifer (2008) can come also from the worker characteristics and not only from the type of firms or jobs in each sector. With our approach we are able to discuss the wage differential across sectors for a given worker and job. On the other hand, the mapping of our estimates to productivity levels depends on the structure of the labor and goods market, because we rely on data on wages rather than quantities produced or profits of the firm; as those other

<sup>16</sup>58% of the workers are part of the shadow economy according to our estimates.

<sup>17</sup>Households are explicitly guaranteed that their answers have no legal implications and cannot be used against them by any government agency.

studies do. For the purposes of this paper this is not important since our object of interest is the income of the worker in each sector. Our results can shed light on the productive structure of the two sectors once the link between wages and productivity is specified.<sup>18</sup>

The remaining of this section is organized as follows: first, we present the data and show how we identify informal workers. Second, the empirical specification is presented and last, the results are shown and discussed.

## 4.1 Data

Our source of information is the household survey (ECH by the Spanish acronym) collected on a monthly basis by the official statistical agency in Colombia (DANE). Our sample is for the year 2013 and comprises 170.000 observations of workers. The sample includes personal information such as age, gender, years of education and also labor market related variables including hours worked, number of jobs, type of job, income sources and social security affiliation. All of the information is self-reported by the worker.

The variables we use from the survey can be grouped into 4 categories: worker characteristics, job characteristics, worker-firm relationship and social security status. A linear combination of the variables in the first three categories is used to construct a factor that captures the variability of wages. The fourth category is used to classify individuals as formal or informal workers. Below we provide a brief description of the variables included in each category, for more detailed information see Appendix B.

**Worker characteristics** capture the type of worker. They include: age, gender, education level and work experience in previous jobs.

**Job characteristics** describe the type of job and task that the worker does. The variables included are: number of workers in the firm (size), industry to which the firm belongs, geographical location of the firm and the task the worker has to do.

**Worker-firm relationship** involves the information about the type of contract and the wage determination. The variables included here are: The wage of the worker, number of working hours, the length of the match, whether the worker is hired through an intermediary firm and whether the worker belongs to a union.

**Social security status** determines whether the worker is affiliated to social security in its different dimensions, and the type of affiliation. The variables included are: affiliation to the health system, the pension system and the labor accidents insurance, as well as who pays for the affiliation to each component.

---

<sup>18</sup>For example, if it is assumed that there is perfect competition on the labor market, then our measure corresponds directly to the worker's marginal productivity. With the additional assumption of a production function with constant returns to scale, our measure also reflects the average productivity of the worker.

## Classification of workers into formal and shadow workers

Colombian regulation provides for labour tax payments (payroll taxes) and the affiliation to social security to be done jointly. Therefore, the affiliation status to the social security system reveals whether the worker's income is taxed and observed by the government, or shadow. We identify a formal worker as a worker affiliated through his own job to all three main components of labor protection: the health security system, the pension system and the accidents insurance policy. With this criteria we estimate that around 58% of the Colombian workers operate in the shadow sector.

When identifying the sector to which the worker belongs we can incur in type I and type II errors, which are respectively: to classify a worker as shadow when he is formal; and to classify a worker as formal when he is shadow. The type I error is not relevant as the affiliation to the social security system is itself a tax on workers, so any worker not affiliated to the system is by definition avoiding labor taxes. On the other hand, there could be shadow workers that decide to register to social security and pay the corresponding contributions, since the affiliation through the alternative subsidized system is mean-tested<sup>19</sup> and they might be not eligible. The incentive for a shadow worker to register and pay is therefore being covered by the health insurance. On the other hand, what induces these workers to remain shadow and misreport their income is paying a lower social contribution and a consequently lower payroll and income tax. We find that by applying the more stringent criterion that requires affiliation not only to the health but also to the pension system and the accidents insurance policy we are able to mitigate the possibility of identifying a shadow worker that registers to social security as a formal worker, as observations with large deviations between the statutory contributions and the actual contributions tend to be for workers that were only affiliated to one or two of the social security provisions (primarily health) but typically not to the accidents insurance.

Finally we could also face the case of a formal worker paying all contributions to the social security system (and being thus classified as formal) but hiding from the government part of his income. This type of worker does pay taxes, but pays less than the amount imposed by the statutory tax imposes. In the case of employees this possibility is mitigated, due to the fact that the firm or the employer are third parties reporting the worker's income and paying the corresponding taxes to the government.<sup>20</sup> The self-employed workers active in the formal sector are also constrained in their income misreporting, since their contractors are the third party in charge to pay the *honorary* tax to tax authorities belong to the formal sector. In conclusion, we believe that these features of the Colombian employment reality allow us to follow the structure of the model by defining tax evasion as working in the shadow economy, while setting aside the aspect of hiding fractions of formal labor income.

---

<sup>19</sup>The housing quality of the recipient is also considered as a criterion to be enrolled of the subsidized system

<sup>20</sup>See for example Kleven, Kreiner, and Saez (2015) for an exploration of the agency role of firms for the implementation of labor taxes and a discussion of the greater tax enforcement when there is third party reporting.

## Colombian labor tax scheme

The main components of the Colombian tax/transfers scheme associated with formal labor income are income taxes, social insurance (payroll) taxes and transfers. First we describe the individual income tax, then the payroll taxes and then the transfers and subsidies. Using this tax scheme we proceed to compute the pre-tax income from the reported income by households and consequently the effective tax rates.

The individual income tax is a progressive tax payable once per year over the total income of one calendar year. The tax is determined by income brackets, and within each of them a fixed amount is paid. The first bracket on which the tax is different from zero starts at 22,219 dollars (annual income in 2013 dollars). The tax rate is increasing across brackets and at the last bracket it reaches 27%.

The social insurance taxes are the payroll tax and the health system contribution. For the case of employees these taxes are paid jointly with the employer; each of the two parties paying a specified fraction. The sum of both (irrespective of who is in charge of making the payment) corresponds to a flat tax rate of 22%.

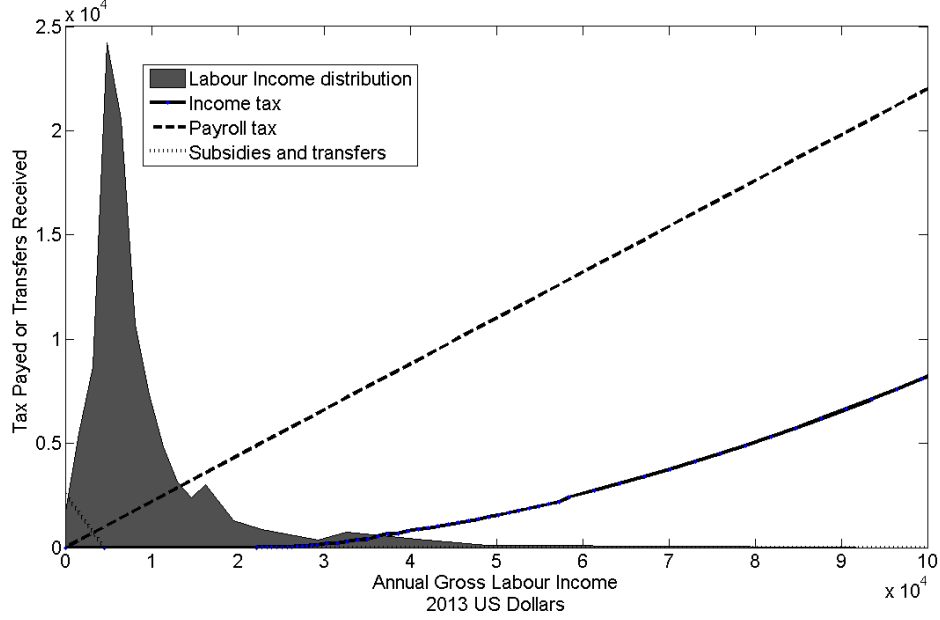
Finally, the bulk of welfare transfers and subsidies in Colombia are granted according to a centralized system that assigns to each household registered in the system a certain score on an index which evaluates needs, life standards, and economic status. The index ranges from 0 to 100, and a series of different welfare programs use it to assign subsidies and transfers, each one according to its own threshold. Part of the questionnaire used to compile the index refers to income of the household. Households have the incentives to misreport income, shadow workers can potentially misreport income while formal workers can be spotted by the system as the reports are crosschecked with the government tax agency. We take an average household that belongs to the subsidized system (meaning the index score is low) (SISBEN) and compute the total transfers it is entitled on that year by the main social programs available. We calculate that those transfers for a household with no formal income could be as large as 2000 dollars per year and reduce to zero for an average household with a full time formal job.

Figure 7 presents the tax scheme decomposed in the three elements discussed and the pre-tax income distribution recovered from reported income and the tax scheme. We see that transfers are an important source of income for the poorer households and that the income tax affects a small fraction of total households.

We have focused on the taxes directly associated with labor income. We do not consider, as they are not part of the instruments we consider in the model, the excise taxes and the corporate income taxes (or taxes over capital gains). If we take that excise taxes are only charged over goods produced in the formal sector and that firms in the formal and shadow economy compete for the same markets then we have that the tax will completely fall on the worker of the formal economy. We leave for further research the possibility of using excise taxes in a setup where the link between goods taxation and labor income has more structure to be analyzed. With our approach we focus



Figure 7: The Colombian Labor Tax Scheme



exclusively on the taxes and transfers that have a direct link with labor income.

### Measuring Income and Wages

Our analysis assumes that all payroll taxes and social security contributions irrespectively of who is administratively charged for the tax are a burden on the worker income. A labor tax that has to be paid by the employer is assumed to be translated in a lower wage for the worker.<sup>21</sup> The workers report their monthly income and the hours worked. To this reported income we input payments that formal workers are entitled to but which are done in a different frequency and are not recorded for the month the survey was conducted. Furthermore, note that we do not include the pension and unemployment insurance contributions as part of the tax burden but we do include them as part of the total income of the worker.

The hourly wage is computed then as the total income divided by the numbers of hours worked. If the worker is a shadow worker we denote it by  $\tilde{w}_i^s$  and if it is formal then is denoted by  $\tilde{w}_i^f$ . These is the key variable that we are going to map to the productivity levels  $w_i^s$  and  $w_i^f$  described in the model.

<sup>21</sup>This is a standard assumption for pretax income computations. The Congressional Budget Office in the US uses the same assumption to compute the effective tax rates.

## 4.2 Empirical specification

The logarithm of both productivities ( $w_i^f$  and  $w_i^s$ ) can be written as a function of a single factor  $F_i$  as follows

$$\log(w_i^f) = \gamma_0^f + \gamma_1^f F_i \quad (39)$$

$$\log(w_i^s) = \gamma_0^s + \gamma_1^s F_i \quad (40)$$

where  $\gamma_0^j, \gamma_1^j$  characterize the linear function in sector  $j \in \{f, s\}$ . We set  $\gamma_1^f = 1$  without loss of generality, given that this will just rescale the factor. The factor is a linear combination of a set of  $n$  variables contained in vector  $X_i$  with weights given by the vector  $\beta$ . Then we have that

$$F_i = \beta X_i \quad (41)$$

The proxy we have for the model productivities are the wages of workers  $\tilde{w}_i^j$  in each sector  $j$ , then we have that<sup>22</sup>

$$\log(\tilde{w}_i^f) = \log(w_i^f) + u_i^f \quad (42)$$

$$\log(\tilde{w}_i^s) = \log(w_i^s) + u_i^s \quad (43)$$

where  $u_i^f$  and  $u_i^s$  are random variables with mean zero. Wages are drawn from a probability distribution where the key location parameters are  $w_i^f$  and  $w_i^s$ , the theoretical concepts in our analysis. In the theoretical analysis we abstract from the underlying variance of the distribution and focus on the limit when it tends to zero. The model is a static economy so we are not concerned with short term variations of wages but rather on the distribution of the location parameters across the population.

Combining equations (39) to (43) we get the specification of the empirical model that corresponds to

$$\log(\tilde{w}_i) = \gamma_0^f + I_i(\gamma_0^s - \gamma_0^f) + (1 + I_i(\gamma_1^s - 1))\beta X_i + u_i \quad (44)$$

where  $I_i$  is an indicator function that takes the value of 1 if type  $i$  works in the shadow economy and  $u_i = I_i u_i^s + u_i^f$ . We estimate (44) by non-linear least squares.

### Ordering of agents and estimated productivities

Note the estimate of parameter  $a$  as  $\hat{a}$ . We proceed to order the individuals in our sample with indexes  $i \in [0, 1]$  such that  $i < i' \iff \hat{\beta} X_i < \hat{\beta} X_{i'}$ . We compute the index of each individual

---

<sup>22</sup>Note that, as discussed earlier,  $w_i^j$  is only observed if type  $i$  works in sector  $j$ .

using the following formula

$$i = \frac{\hat{\beta}X_i - \min_{i'}\{\hat{\beta}X_{i'}\}}{\max_{i'}\{\hat{\beta}X_{i'}\}}$$

that is just rescaling the factor using the minimum and the maximum values it takes in the sample. The estimated productivities of each type  $i$  then correspond to

$$\hat{w}_i^f = \exp\left\{\hat{\gamma}_0^f + \hat{\beta}X_i\right\} \quad (45)$$

$$\hat{w}_i^s = \exp\left\{\hat{\gamma}_0^s + \hat{\gamma}_1^s \hat{\beta}X_i\right\}. \quad (46)$$

### Single-crossing condition

The single-crossing condition states that the ratio  $w_i^f/w_i^s$  is increasing in type. Using (45) and (46) this ratio can be written as

$$\frac{\hat{w}_i^f}{\hat{w}_i^s} = \exp\left\{\hat{\gamma}_0^f - \hat{\gamma}_0^s\right\} \exp\left\{(1 - \hat{\gamma}_1^s) \hat{\beta}X_i\right\}$$

Then, if  $\hat{\gamma}_1^s < 1$  holds, the single-crossing condition is satisfied. Recall that we standardized to 1 the marginal (percentile) increase of formal productivity to a marginal increase in the factor. Therefore, this condition states that a marginal increase in the factor has to imply a lower marginal increase in shadow than in formal productivity.

### Top income earners

We standardized the time available for labor in a year equal to 1 and therefore we can interpret  $\tilde{w}_i^j$  as the income of worker  $i$  for full time work at sector  $j$ , then  $\hat{w}_i^f$  corresponds (on average) to the maximum income that type  $i$  can achieve. Nevertheless, some income observations are above the maximum value implied by the factor for the most productive worker working full time. That is, there could be labor income observations  $y_i$  that satisfy

$$y_i > \max_{i'}\{\hat{w}_{i'}^f\} = \hat{w}_1^f \quad (47)$$

We classify the individuals that satisfy this criterion as top earners. These are individuals with a very large wage premium that cannot be accounted for with our benchmark specification and for which the wage does not seem to have the same relationship with the factor as for the rest of the population.

To characterize with more accuracy this behavior at the top of the income distribution we estimate the upper tail of the productivity distribution by fitting a Type I Pareto distribution for the gross wage  $\tilde{w}$  of top earners. The support of the distribution is given by  $[\hat{w}_1^f, \infty)$  and the shape parameter is estimated by maximum likelihood.

A final adjustment has to be made to the index of agents. To fit the top earners in the type space  $[0, 1]$  we compress the indexes on non-top earners to the interval  $[0, k]$  and top earners are assigned to  $[k, 1]$  and ordered by their gross wage.

### Distribution of types

The assignment of indexes for each observation and their corresponding sampling weights implies a discrete distribution of workers (non-top earners). The continuous distribution of types is obtained by a kernel density estimation with a linear interpolation at the evaluation points. The estimated kernel distribution gives us the distribution of types in the interval  $[0, k]$ .

For top earners we have a Pareto distribution for productivities with the support  $[\max_i \{\hat{w}_{it}^f\}, \infty)$  but this distribution can be replicated by different types distributions in  $[k, 1]$  at the types space, provided that the formal productivities  $w_i^f$  for  $i \in [k, 1]$  are adjusted accordingly. This phenomenon does not occur with non-top earners because their productivity profiles are given by our parametric model.

There are two requirements that the distribution of types and productivity profiles of top earners satisfy always: the total mass of the distribution has to coincide with the mass of top earners and that  $\lim_{i \rightarrow 1} w_i^f = \infty$ .

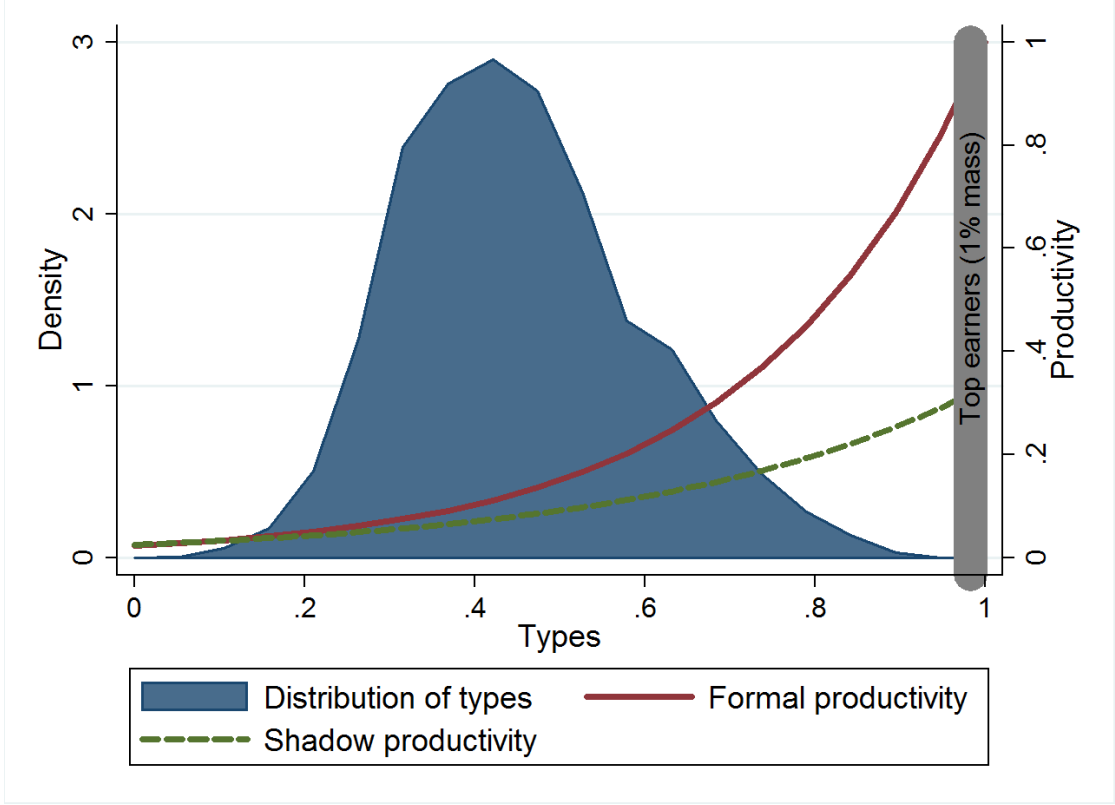
### 4.3 Estimation results

Here we discuss the results of the estimation of the formal productivity ( $w_i^f$ ), the informal productivity ( $w_i^s$ ) and the distribution of types ( $\mu_i$ ). Parameter estimates for  $\beta$  and the detailed description of the variables included in  $X_i$  are presented in Appendix B.

Figure 8 presents the estimated productivities and the types distribution for non-top earners. The estimated values of  $\gamma_0^f$  and  $\gamma_0^s$  are almost identical with  $\hat{\gamma}_0^s$  slightly greater so type 0 is slightly more productive in the shadow economy. The single-crossing condition is supported by the data since the hypothesis  $\gamma_1^s < 1$  is not rejected at a 1% confidence level. The most productive individual among non-top earners is almost three times more productive in the formal economy than in the shadow economy.

Top earners are assigned to the set  $[0.98, 1]$ , the estimated value of the shape parameter of the Pareto distribution is 1.81 and comprise a mass of about 1% of the total population (details of the estimation are presented in Appendix B). The shaded region in Figure 8 corresponds to the top earners. We do not plot their productivity profiles and density. Recall that what is identified is the distribution of formal productivities at the top with support  $[\max_i \{\hat{w}_{it}^f\}, \infty)$  and this can be matched with many different combinations of formal productivity and probability density specifications in the types space; all of them equivalent for the optimal taxation problem that solves the planner. We assume that the relation between the shadow and the formal productivity from the main part of the distribution of types holds also for the top earners.

Figure 8: Estimated productivities and types distributions



## 5 Calibrated exercise

Given the productivity schedules estimated in the previous section, we calibrate the utility function and derive the optimal allocations for Colombia.

### 5.1 Calibration of the utility function

We assume that the agents' utility function is

$$U(c, n) = \log \left( c - \Gamma \frac{n^{1+\frac{1}{\zeta}}}{1+\frac{1}{\zeta}} \right), \quad n \in [0, 1]. \quad (48)$$

The parameter  $\zeta$  is the elasticity of labor supply. Since we consider a permanent tax reform, the relevant notion is the steady-state intensive margin elasticity. We fix  $\zeta$  at different values and find  $\Gamma$  which minimizes the deviation of selected  $K$  model moments  $(m_k^{model}(\zeta, \Gamma))_{k=1}^K$  from the

corresponding data moments  $(m_k^{data})_{k=1}^K$  according to the loss function

$$L(\zeta, \Gamma) = \sum_{k=1}^K \left( \frac{m_k^{model}(\zeta, \Gamma) - m_k^{data}}{m_k^{data}} \right)^2. \quad (49)$$

We use three moments: the share of shadow workers in total employment, the share of shadow income in total income and the average total income. The first two moments capture the relative size of the shadow economy, while the third one controls for the total production of Colombia. Chetty, Guren, Manoli, and Weber (2011) recommend using the steady-state intensive elasticity of 0.33, which we treat as a benchmark. However, the estimates behind this number implicitly incorporate responses on multiple margins, possibly also shifting labor to the shadow economy. Since we model this response explicitly, the correct value of elasticity could lower. Hence, we consider also the values of 0.2 and 0.1. Table 1 shows the matched moments for different values of the elasticity of labor supply.

Table 1

Moments	Actual economy	Model economy for different values of elasticity $\zeta$		
		$\zeta = 0.33$	$\zeta = 0.2$	$\zeta = 0.1$
share of shadow workers	57.99%	64.51%	62.12%	60.53%
share of shadow income	30.94%	23.25%	25.24%	26.64%
mean total income [USD]	7166	6673	6659	6677

The model replicates well the magnitude of the shadow economy for a range of elasticities of labor supply. We conclude that the empirical distribution of productivities and the actual tax schedule can explain the high level of informality in Colombia.

## 5.2 Optimal allocations

We find the optimum for the two social welfare functions. First, we use the Rawlsian welfare criterion, which puts all the weight on the individual with the lowest utility level. Since both formal and shadow productivities are increasing with type, the Rawlsian planner cares only about the lowest type. Second, we derive the Utilitarian optimum with the planner that maximizes the average utility level in the economy. In each case we require that the planner obtains the same net tax revenue as the actual tax schedule.

The optimal allocations are described in Table 2. The Rawlsian planner would displace close to 22% of the workforce to informality. The share of shadow income falls even more, since only the least productive workers end up in the shadow economy. The Utilitarian planner would cut the size of the informal sector even more, to less than 1%. The Utilitarian planner cares mainly about

workers in the middle of the distribution, where the density of types is high. Hence, this planner is not willing to set high marginal tax rates at the bottom, as it would reduce the utility of the workers in the middle. As the tax rate at the bottom is low, few workers are displaced to the shadow economy.

The welfare gains from implementing the optimum are large. The Rawlsian planner manages to increase the transfers to the workers with no formal income by 85% in comparison to the actual tax and transfer system. It translates into welfare gains of 40% to 50% in consumption equivalent terms. The Utilitarian planner takes into consideration the welfare cost of increased taxation of the high types and expands the redistribution less. Nevertheless, the transfers received by the bottom types increase by more than 55% in comparison to the actual tax system in Colombia and welfare gains are close to of 20% in terms of consumption. In order to make sure that the welfare gains are not driven by a thick Pareto tail at the top, we recompute the optima without the top tail (see the last row of Table 2).<sup>23</sup> The welfare gains are naturally smaller, since the top earners constitute a sizable source of tax revenue. However, it is clear that most of the welfare gains come from the efficient taxation of the ordinary workers and not from the very rich.

Table 2

Moments	Actual economy	Optimal Rawlsian allocation			Optimal Utilitarian allocation		
		$\zeta = 0.33$	$\zeta = 0.2$	$\zeta = 0.1$	$\zeta = 0.33$	$\zeta = 0.2$	$\zeta = 0.1$
share of shadow workers	57.99%	21.68%	21.68%	21.68%	0.17%	0.18%	0.19%
share of shadow income	30.94%	5.59%	6.33%	6.98%	0.02%	0.03%	0.03%
mean total income [USD]	7165	6671	6967	7112	6825	7086	7245
welfare (cons. equiv.)	100%	151.8%	147.8%	142%	121.3%	120.9%	119.7%
welfare w/o top tail (cons. equiv.)	100%	136.5%	135%	133.6%	116.8%	117%	117.4%

Figure 9 demonstrates how the optimal tax schedule is determined. Recall that the shadow economy imposes an upper bound on the tax rate. If the tax rate of type  $i$  exceeds  $1 - w_i^s/w_i^f$ , the return to shadow labor is strictly greater than the return to formal labor. No agent of type  $i$  would be willing to supply formal labor at such terms. As is evident from the figure, all bottom types face tax rate above the upper bound. Hence, they are bunched together at the zero formal income. From equation (37) we know that workers who are not bunched face the marginal tax rate that is a minimum of the two expressions: the standard Mirrleesian tax rate given by a Diamond (1998) and the upper bound  $1 - w_i^s/w_i^f$ . In all our calibrations the upper bound plays a dominant role (see Figure 9). For the Utilitarian planner with elasticity of 0.33 the standard Mirrleesian tax rate dives under the upper bound just for some high types. For the Rawlsian planner, as well as in the cases of lower elasticity of labor supply, the Mirrleesian tax rate does not intercept the upper bound below

<sup>23</sup>In this case the distribution of types has finite support. The mass of the excluded tail is 0.0045.

the upper tail and hence does not influence the optimal tax in the main part of distribution. In contrast, in all our calibrations some of the upper tail workers are taxed according to the Diamond (1998) formula (the upper tail is not represented on Figure 9). We conclude that the optimal tax schedule of workers below the upper tail is predominantly determined by the shadow economy considerations. However, the usual labor supply responses are important for taxing very productive workers.

Figure 9: The role of the upper bound

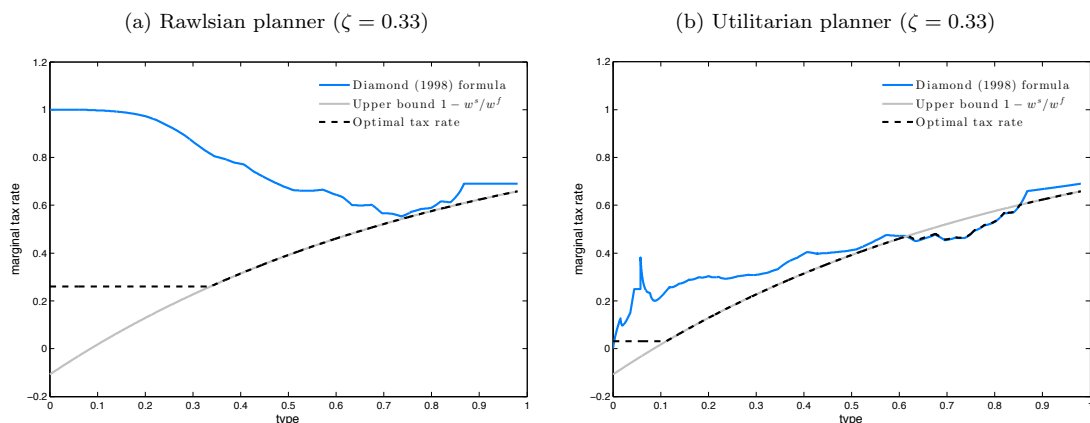


Figure 9 informs us also what would happen if the shadow economy was neglected and the standard Mirrleesian tax was implemented. All the types for which the tax rate exceeds the upper bound would be displaced to the shadow economy. Moreover, many types for which the Mirrleesian tax rate is below the upper bound are likely to move to the shadows as well.<sup>24</sup> Hence, the implementation of the usual tax formula which does not account for the shadow economy would lead to a dramatic fall in tax revenue.

How does the optimal tax schedule compares with the one implemented at the time in Colombia? The actual tax schedule involves high 45% marginal rate at low levels of income, implied by phasing-out of transfers (see Figure 10). As income increases the rate drops to 22% and remains flat - workers with this income pay only the flat payroll tax. The progressive income tax starts at the high income level and gradually increases the marginal tax, reaching 49% for the top earners (at income levels not represented at Figure 10)).<sup>25</sup> In comparison to the actual tax rate, the optimal tax rates are lower at low levels of income and much higher elsewhere. Lower marginal rates at the bottom mean transfers are phased-out more slowly, so less productive workers have less incentives to move to the informal sector. Higher marginal tax rates elsewhere imply that the richest agents pay much higher total tax than in the actual economy, which allows the planner to finance the generous transfer

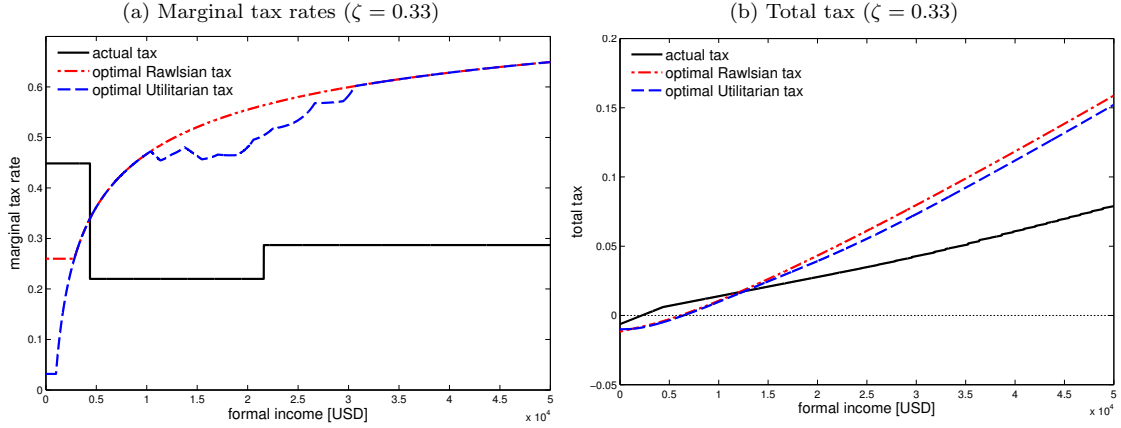
<sup>24</sup>The tax burden accumulated at the low income levels is likely to outweigh the gain from higher return to formal labor at the high income levels.

<sup>25</sup>The progressive tax is a step function with more than 80 steps of varying width and Figure 10 (a) shows its smoothly approximation. The true tax involves 0 rate at the interior of each step and an unbounded rate between steps, hence it cannot be represented on such graph.



(Figure 10 (b)). The tax rates at lower elasticities have very similar shape, as they are determined by the upper bound.

Figure 10: The optimal tax schedule



## 6 Conclusions

A large fraction of the economic activity in most countries is informal. This paper incorporates this fact into the optimal income tax theory. We find that the shadow economy puts severe restrictions on the taxes the government can levy, often leading to a welfare loss. However, in some cases the shadow economy can raise welfare by improving both redistribution and efficiency. If the informal sector suppresses productivity differences between workers, the government can tax high earners more when the low productivity workers are employed informally. Furthermore, the shadow economy shelters poor workers from distortions implied by the taxation of the rich, allowing for more efficient allocation of labor.

The mechanism proposed has a quantitatively sizable effect. In the case of Colombia, the government that cares only about the poor would optimally choose to have 22% of workers in the shadow economy. Nevertheless, the observed levels of informality are much higher than that. According to our model, the large size of the Colombian shadow economy is explained by high marginal tax rates at low levels of income. The optimal tax schedule features lower rates at the bottom, leading to a smaller informal sector, and higher rates above, raising more revenue from top earners.

This paper suggests that allowing less productive people to collect welfare benefits and simultaneously work in the shadow economy could be desirable. Moreover, policies designed to deter the creation of informal jobs should focus on the jobs taken by the workers with the potential for high formal earnings. It is important to stress that the way the shadow economy is modeled in this paper abstracts from many issues, such as competition between formal and informal firms, lack of regulation and law enforcement, as well as potential negative externalities caused by the informal

activity. All those phenomena are likely to reduce the potential welfare gains from exploiting the shadow economy.

## References

- ALBRECHT, J., L. NAVARRO, AND S. VROMAN (2009): “The Effects of Labour Market Policies in an Economy with an Informal Sector,” *The Economic Journal*, 119(539), 1105–1129.
- ALLINGHAM, M. G., AND A. SANDMO (1972): “Income tax evasion: A theoretical analysis,” *Journal of Public Economics*, 1(3), 323–338.
- ALVAREZ-PARRA, F., AND J. M. SÁNCHEZ (2009): “Unemployment insurance with a hidden labor market,” *Journal of Monetary Economics*, 56(7), 954–967.
- AMARAL, P. S., AND E. QUINTIN (2006): “A competitive model of the informal sector,” *Journal of Monetary Economics*, 53(7), 1541–1553.
- BOADWAY, R., K. CUFF, AND M. MARCHAND (2000): “Optimal Income Taxation With Quasi-Linear Preferences Revisited,” *Journal of Public Economic Theory*, 2(4), 435–460.
- BRENDON, C. (2013): “Efficiency, equity, and optimal income taxation,” .
- CHETTY, R., A. GUREN, D. MANOLI, AND A. WEBER (2011): “Are micro and macro labor supply elasticities consistent? A review of evidence on the intensive and extensive margins,” *The American Economic Review*, 101(3), 471–475.
- CHONÉ, P., AND G. LAROQUE (2010): “Negative marginal tax rates and heterogeneity,” *The American Economic Review*, pp. 2532–2547.
- DE SOTO, H. (1990): “The other path: The invisible revolution in the third world,” .
- (2000): *The mystery of capital: Why capitalism triumphs in the West and fails everywhere else*. Basic books.
- DIAMOND, P. A. (1998): “Optimal income taxation: an example with a U-shaped pattern of optimal marginal tax rates,” *American Economic Review*, pp. 83–95.
- EBERT, U. (1992): “A reexamination of the optimal nonlinear income tax,” *Journal of Public Economics*, 49(1), 47–73.
- FRÍAS, J. A., T. KUMLER, AND E. A. VERHOOGEN (2013): “Enlisting Employees in Improving Payroll-Tax Compliance: Evidence from Mexico,” .
- GOMES, R. D., J.-M. LOZACHMEUR, AND A. PAVAN (2014): “Differential taxation and occupational choice,” .

- JUTTING, J., J. R. D. LAIGLESIA, ET AL. (2009): “Is informal normal?: towards more and better jobs in developing countries,” .
- KLEVEN, H. J., C. T. KREINER, AND E. SAEZ (2015): “Why Can Modern Governments Tax So Much? An Agency Model of Firms as Fiscal Intermediaries,” .
- KOPCZUK, W. (2001): “Redistribution when avoidance behavior is heterogeneous,” *Journal of Public Economics*, 81(1), 51–71.
- LA PORTA, R., AND A. SHLEIFER (2008): “The unofficial economy and economic development,” Discussion paper, National Bureau of Economic Research.
- MEGHIR, C., R. NARITA, AND J.-M. ROBIN (2015): “Wages and Informality in Developing Countries,” *American Economic Review*, 105(4), 1509–46.
- MILGROM, P., AND I. SEGAL (2002): “Envelope theorems for arbitrary choice sets,” *Econometrica*, pp. 583–601.
- MIRPLEES, J. A. (1971): “An exploration in the theory of optimum income taxation,” *The Review of Economic Studies*, pp. 175–208.
- MUSSA, M., AND S. ROSEN (1978): “Monopoly and product quality,” *Journal of Economic Theory*, 18(2), 301–317.
- PIKETTY, T., AND E. SAEZ (2013): “Optimal Labor Income Taxation,” *Handbook of Public Economics*, 5, 391.
- RAUCH, J. E. (1991): “Modelling the informal sector formally,” *Journal of Development Economics*, 35(1), 33–47.
- SAEZ, E. (2001): “Using elasticities to derive optimal income tax rates,” *The Review of Economic Studies*, 68(1), 205–229.
- SCHNEIDER, F., A. BUEHN, AND C. E. MONTENEGRO (2011): “Shadow Economies all over the World: New Estimates for 162 Countries from 1999 to 2007,” *Handbook on the shadow economy*, pp. 9–77.
- WASEEM, M. (2013): “Taxes, Informality and Income Shifting: Evidence from a Recent Pakistani Tax Reform,” *Available at SSRN 2388116*.

## A Proofs from Section 2

**Proof of Proposition 1.** Omitted.  $\square$

**Proof of Proposition 2.** Note that the first-best allocation is consistent with the additional constraint (5), hence it is the solution to the planner's problem. Essentially, conditional on truthfully revealing type, incentives of the agent and the planner regarding the shadow labor are perfectly aligned. If a given type pays taxes according to the true type, choosing shadow labor in order to maximize utility cannot hurt the social welfare.  $\square$

**Proof of Proposition 3.** In the first-best,  $U(c_L, n_L) \geq U(c_H, n_H)$ . By assumption of  $v'(0) = 0$ , we know that  $n_L^f > 0$ . Then the utility of  $H$  mimicking  $L$  is  $U\left(c_L, \frac{w_L^f}{w_H^f} n_L^f\right) > U(c_L, n_L^f) \geq U(c_H, n_H)$ , which violates  $IC_{H,L}$ . Hence, the optimum is not the first-best.

Suppose that at the optimum  $IC_{H,L}$  does not bind. First, let's consider the case in which  $U(c_H, n_H) > U(c_L, n_L)$ . Since  $IC_{H,L}$  is slack, the planner may increase transfers from  $H$  to  $L$ , which raises welfare, so it could not be the optimum in the first place. Second, suppose that  $U(c_L, n_L) \geq U(c_H, n_H)$ . It can happen only if  $n_L^s > 0$ . Otherwise, as we have shown above,  $IC_{H,L}$  is violated. If  $n_L^s > 0$  and  $IC_{H,L}$  is slack, the planner can marginally decrease  $n_L^s$  and increase  $n_L^f$ , which generates free resources. Hence, at the optimum  $IC_{H,L}$  has to bind.

Suppose that  $IC_{L,H}$  binds. If the resource constraint is satisfied as equality, it may happen only if  $L$  type is paying a positive tax, while  $H$  type receives a transfer. Then the planner can improve welfare by canceling the redistribution altogether and reverting to *laissez-faire*, where none of the incentive constraints bind.  $\square$

**Lemma A.1.** *At the optimum either  $U(c_L, n_L) = U(c_H, n_H)$  and  $n_L^s > 0$ , or the following optimality condition holds*

$$\min \left\{ \frac{v'(n_L)}{w_L^f} - \left( \mu_L + \mu_H \frac{v'(n_{H,L})}{w_H^f} \right), n_L^f \right\} = 0, \quad (50)$$

where  $n_{i,-i} = \frac{w_{-i}^f}{w_i^f} n_{-i}^f + n_i^s \left( \frac{w_{-i}^f}{w_i^f} n_{-i}^f \right)$  is the total labor supply of type  $i$  pretending to be of type  $-i$ .

Suppose that  $v''$  is nondecreasing. If  $\frac{w_H^f}{w_L^f} g(w_H^s) \geq g(w_L^s)$  then this optimality condition is sufficient for the optimum.

**Proof of Lemma A.1.** If  $U(c_L, n_L) = U(c_H, n_H)$  and  $n_L^s = 0$ , then such allocation is not incentive compatible. The proof is identical as the proof of the claim that the first-best is not incentive compatible in Proposition 3. Hence, if  $U(c_L, n_L) = U(c_H, n_H)$ , then  $n_L^s > 0$ .

Let's consider the case in which  $U(c_H, n_H)$  is always greater than  $U(c_L, n_L)$ .  $IC_{H,L}$  has to bind, otherwise the planner could equalize utilities of both types. Consider changing  $n_L^f$  by a small amount and adjusting  $T_L$  such that  $IC_{H,L}$  is satisfied as equality. It means that

$$\frac{dT_L}{dn_L^f} = w_L^f \mu_H \left( 1 - \frac{v'(n_{H,L})}{w_H^f} \right).$$

This perturbation affects social welfare by

$$\frac{dU(c_L, n_L)}{dn_L^f} = w_L^f - \frac{\partial T_L}{\partial n_L^f} - v'(n_L) = w_L^f \left( \mu_L + \mu_H \frac{v'(n_{H,L})}{w_H^f} \right) - v'(n_L).$$

Optimum requires that either  $\frac{dU(c_L, n_L)}{dn_L^f} = 0$  or  $\frac{dU(c_L, n_L)}{dn_L^f} < 0$  and  $n_L^f = 0$ , which results in (50). Sufficiency of this first order condition depends on the second order derivative of welfare with respect to the perturbation. In order to have the second derivative well behaved, we are going to assume that  $v''$  is nondecreasing. Then, we need to consider two cases (see Table 3). If  $\frac{w_H^f}{w_L^f} g(w_H^s) \geq g(w_L^s)$  holds, then  $\frac{dU(c_L, n_L)}{dn_L^f}$  is non-increasing in  $n_L^f$ . It means that the optimality condition (50) is sufficient. If  $\frac{w_H^f}{w_L^f} g(w_H^s) < g(w_L^s)$ , then  $\frac{dU(c_L, n_L)}{dn_L^f}$  is not monotone in  $n_L^f$  and it may be the case that (50) points at either local maximum which is not a global maximum, or at the local minimum.

Table 3: Second order derivative of welfare with respect to the perturbation

The case of $\frac{w_H^f}{w_L^f} g(w_H^s) \geq g(w_L^s)$			
	$n_L^f < g(w_L^s)$	$g(w_L^s) < n_L^f < \frac{w_H^f}{w_L^f} g(w_H^s)$	$\frac{w_H^f}{w_L^f} g(w_H^s) < n_L^f$
$\frac{d^2 U(c_L, n_L)}{dn_L^{f2}} =$	0	$-v''(n_L^f) < 0$	$\mu_H \left( \frac{w_H^f}{w_L^f} \right)^2 v'' \left( \frac{w_H^f}{w_L^f} n_L^f \right) - v''(n_L^f) < 0$
The case of $\frac{w_H^f}{w_L^f} g(w_H^s) < g(w_L^s)$			
	$n_L^f < \frac{w_H^f}{w_L^f} g(w_H^s)$	$\frac{w_H^f}{w_L^f} g(w_H^s) < n_L^f < g(w_L^s)$	$g(w_L^s) < n_L^f$
$\frac{d^2 U(c_L, n_L)}{dn_L^{f2}} =$	0	$\mu_H \left( \frac{w_H^f}{w_L^f} \right)^2 v'' \left( \frac{w_H^f}{w_L^f} n_L^f \right) > 0$	$\mu_H \left( \frac{w_H^f}{w_L^f} \right)^2 v'' \left( \frac{w_H^f}{w_L^f} n_L^f \right) - v''(n_L^f) < 0$

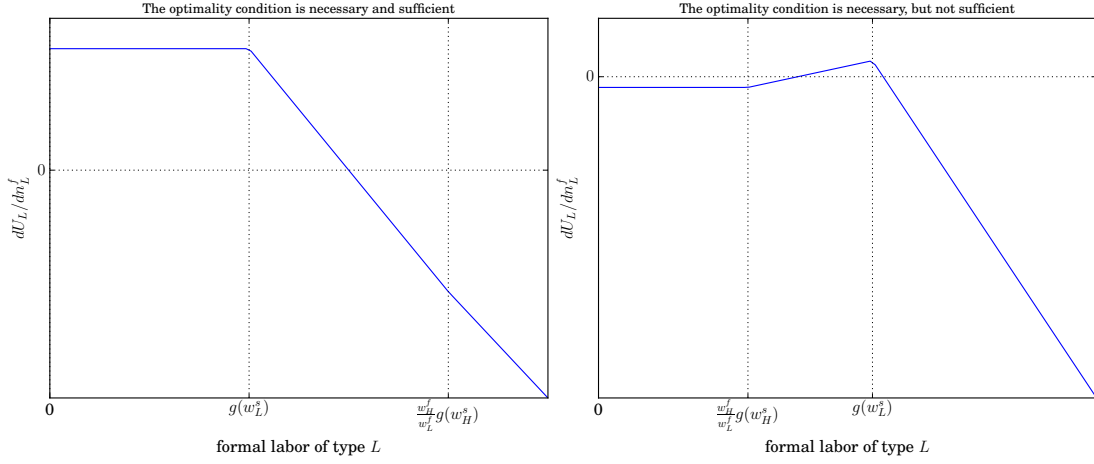
Figure 11 shows these two cases. In the first panel  $\frac{w_H^f}{w_L^f} g(w_H^s) \geq g(w_L^s)$  holds and the optimality condition (50) always points at the optimum (in this case, the value of  $n_L^f$  where  $\frac{dU(c_L, n_L)}{dn_L^f} = 0$ ).

In the second panel  $\frac{w_H^f}{w_L^f} g(w_H^s) < g(w_L^s)$  holds and the optimality condition is not sufficient. There are three points that satisfy condition (50): local maximum at  $n_L^f = 0$ , local minimum with  $n_L^f \in \left( \frac{w_H^f}{w_L^f} g(w_H^s), g(w_L^s) \right)$  and the other local maximum with  $n_L^f > g(w_L^s)$ .  $\square$

**Proof of Proposition 4.** In the proof of Lemma A.1 above we described the impact of changing formal labor of  $L$  on the social welfare,  $\frac{dU(c_L, n_L)}{dn_L^f}$ . The condition (10) describes situations when the impact of the perturbation is non-positive at  $n_L^f = 0$ . From Figure 11 it is clear that if it is not the case, type  $L$  will never optimally work in the shadow economy.

Suppose that  $\frac{w_H^f}{w_L^f} g(w_H^s) \geq g(w_L^s)$ . Condition (10) implies that  $\frac{dU(c_L, n_L)}{dn_L^f}$  is always non-positive, so it is optimal to reduce  $n_L^f$  as long as  $U(c_H, n_H) > U(c_L, n_L)$ . From Lemma A.1 we know also that  $U(c_H, n_H) > U(c_L, n_L)$  if  $L$  works only formally, so it is optimal to place type  $L$  in the shadow economy.

Figure 11: Sufficiency of the optimality condition



Now suppose that  $\frac{w_H^f}{w_L^f}g(w_H^s) < g(w_L^s)$ . Condition (11) means that the maximum of  $\frac{dU(c_L, n_L)}{dn_L^f}$  attained at  $n_L^f = g(w_L^s)$  (see Figure 11) is non-positive. Therefore, it is optimal to reduce  $n_L^f$  until utilities of both types are equalized, which can happen only when  $L$  works in the shadow economy. Condition (11) is sufficient, but not necessary for  $L$  to work in the shadow economy, because the social welfare changes in a non-monotone way with  $n_L^f$ . If (11) is not satisfied, marginally increasing  $n_L^s$  from 0 is bad for welfare, but increasing it further may eventually lead to welfare gains, and the total effect on welfare is ambiguous.  $\square$

**Proof of Proposition 5.** Suppose that optimally  $n_L^s > 0$ . From Figure 11 it is clear that in such situation it is in the best interest of type  $L$  to work exclusively in the shadow economy. However, if  $w_L^s > w_H^s$  and  $n_L^f = 0$ , the incentive compatibility constraint of the type  $H$  implies that

$$U(c_L, n_L) = U(w_L^s n_L^s - T_L, n_L^s) > U(w_H^s n_{H,L}^s - T_L, n_{H,L}^s) = U(c_H, n_H).$$

Since the planner is Rawlsian, such allocation is not desirable. The planner will rather stop decreasing  $n_L^f$  at the point where utilities of both types are equal. On the other hand, if  $w_L^s \leq w_H^s$  then

$$U(c_L, n_L) = U(w_L^s n_L^s - T_L, n_L^s) \leq U(w_H^s n_{H,L}^s - T_L, n_{H,L}^s) = U(c_H, n_H),$$

so the planner will optimally decrease  $n_L^f$  to zero.  $\square$

**Proof of Proposition 6.** In order to examine when the optimum welfare is strictly higher than in the standard Mirrlees model, we will compare utility of type  $L$  in the standard Mirrlees model ( $U(c_L^M, n_L^M)$ ) and in the shadow economy model, when  $L$  is working only in the shadow economy

$(U(c_L^{SE}, n_L^{SE}))$ . Clearly, when the second scenario yields higher utility, the existence of the shadow economy is welfare improving.

In the standard Mirrlees model, the binding constraint is  $U(w_H^f n_H^M, n_H^M) - T_H^M = U(w_L^f n_L^M, \frac{w_L^f}{w_H^f} n_L^M) - T_L^M$ . Together with the resource constraint it means that  $T_L^M = \mu_H \left( U(w_L^f n_L^M, \frac{w_L^f}{w_H^f} n_L^M) - U(w_H^f n_H^M, n_H^M) \right)$ . Now, the utility of type  $L$  is

$$U(c_L^M, n_L^M) = U(w_L^f n_L^M, n_L^M) - T_L = U(w_L^f n_L^M, n_L^M) - \mu_H \left( U(w_L^f n_L^M, \frac{w_L^f}{w_H^f} n_L^M) - U(w_H^f n_H^M, n_H^M) \right).$$

Using the same steps, we can express the utility of type  $L$  working only in the shadow economy as

$$U(c_L^{SE}, n_L^{SE}) = U(w_L^s n_L^{SE}, n_L^{SE}) - \mu_H \left( U(w_H^s n_{H,L}^{SE}, n_{H,L}^{SE}) - U(w_H^f n_H^{SE}, n_H^{SE}) \right).$$

Since there are no distortions at the top and no wealth effects,  $n_H^M = n_H^{SE}$ . The shadow economy is welfare improving,  $U(c_L^{SE}, n_L^{SE}) - U(c_L^M, n_L^M) > 0$ , iff

$$U(w_L^s n_L^{SE}, n_L^{SE}) - U(w_L^f n_L^M, n_L^M) + \mu_H \left( U(w_L^f n_L^M, \frac{w_L^f}{w_H^f} n_L^M) - U(w_H^s n_{H,L}^{SE}, n_{H,L}^{SE}) \right) > 0.$$

The first difference (the efficiency gain) is positive if  $w_L^s > \bar{w}_L^s$ . The second difference (the redistribution gain) is positive when  $w_H^s < \bar{w}_H^s$ . Hence, when both inequalities hold weakly and at least one holds strictly, the existence of the shadow economy improves welfare in comparison to the standard Mirrlees model.

Now we will show that when the inequalities hold in the other direction, the shadow economy hurts welfare. Suppose that  $w_L^s = \bar{w}_L^s$  and  $w_H^s = \bar{w}_H^s$ . We will prove that allocation  $^{SE}$  is a unique optimum at this point. First we will show that when the redistribution gain is non-positive, it is true that  $n_H^{SE} > \frac{w_L^f}{w_H^f} n_L^M$ . Suppose on the contrary that  $n_H^{SE} \leq \frac{w_L^f}{w_H^f} n_L^M$ . Then we can write the following sequence of inequalities

$$U\left(w_L^f n_L^M, \frac{w_L^f}{w_H^f} n_L^M\right) \geq U\left(w_H^f n_H^{SE}, n_H^{SE}\right) > U\left(w_H^s n_H^{SE}, n_H^{SE}\right).$$

The first inequality comes from the fact that  $\frac{w_L^f}{w_H^f} n_L^M$  is below the efficient level of labor supply of type  $H$ , so lowering the labor of this type even further to  $n_H^{SE}$  will decrease the utility. The second inequality is simply implied by our assumption  $w_H^f > w_H^s$ . This sequence of inequalities implies that the redistribution gain is strictly positive. Hence, if the redistribution gain is non-positive,  $n_H^{SE} > \frac{w_L^f}{w_H^f} n_L^M$  holds.

Note that  $n_H^{SE} > \frac{w_L^f}{w_H^f} n_L^M$  means that the optimal allocation of the standard Mirrlees model is not incentive-compatible with the shadow economy - deviating type  $H$  would supply some additional

shadow labor. Hence, any allocation which yields the social welfare equal or higher than  $U(c_L^M, n_L^M)$  has to involve type  $L$  working in the shadow economy.

Let's go back to the optimal allocation with the shadow economy, when  $w_L^s = \bar{w}_L^s$  and  $w_H^s = \bar{w}_H^s$ . From the considerations above we know that the optimum involves some shadow labor. If we sum the efficiency gain and the redistribution gain divided by  $\mu_H$  and rearrange the terms, we get

$$\left( U\left(w_L^f n_L^M, \frac{w_L^f}{w_H^f} n_L^M\right) - U\left(w_L^f n_L^M, n_L^M\right) \right) - (U(w_H^s n_H^{SE}, n_H^{SE}) - U(w_L^s n_L^{SE}, n_L^{SE})) = 0.$$

The expression in the first brackets is positive. Hence, the second brackets are positive as well, which means that  $w_H^s > w_L^s$ . By Proposition 5 type  $L$  will work exclusively in the shadow economy.

To sum up, we know that at  $(w_L^s, w_H^s) = (\bar{w}_L^s, \bar{w}_H^s)$  the optimum of the shadow economy model is unique and involves type  $L$  working entirely in the shadow economy. Consequently, a decrease in the shadow productivity of type  $L$  or an increase in the shadow productivity of type  $H$  leads to a strict welfare loss, since it either decreases the effective productivity of type  $L$  or decreases the transfer type  $L$  receives.  $\square$

**Proof of Proposition 7.** Suppose that  $\lambda_i \leq \lambda_{-i}$ . In this case the  $IC_{i,-i}$  may bind (it will if the inequality is strict), while  $IC_{-i,i}$  is always slack. The planner will not distort the allocation of type  $i$ . Without distortions, this type will never work in the shadow economy.

Suppose that  $\lambda_i > \lambda_{-i}$ , so that  $IC_{-i,i}$  binds. Perturb  $n_i^f$  and adjust  $T_i$  such that  $IC_{-i,i}$  holds as equality:

$$\frac{dT_i}{dn_i^f} = w_i^f \mu_{-i} \left( 1 - \frac{v'(n_{-i,i})}{w_{-i}^f} \right).$$

This perturbation affects social welfare by

$$\begin{aligned} \frac{dW}{dn_i^f} &= \lambda_i \mu_i \left( w_i^f - \frac{\partial T_i}{\partial n_i^f} - v'(n_i) \right) + \lambda_{-i} \mu_{-i} \frac{\mu_i}{\mu_{-i}} \frac{\partial T_i}{\partial n_i^f} \\ &= \lambda_i \mu_i w_i^f \left( \left( 1 - \frac{v'(n_i)}{w_i^f} \right) + \left( \frac{\lambda_{-i}}{\lambda_i} - 1 \right) \mu_{-i} \left( 1 - \frac{v'(n_{-i,i})}{w_{-i}^f} \right) \right) \end{aligned} \quad (51)$$

Suppose that  $\frac{w_{-i}^s}{w_{-i}^f} \geq \frac{w_i^s}{w_i^f}$  and  $n_i^f \leq g(w_i^s)$ , which means that  $v'(n_i) = w_i^s$ . Note that  $\frac{v'(n_{-i,i})}{w_{-i}^f} \geq \frac{w_{-i}^s}{w_{-i}^f} \geq \frac{w_i^s}{w_i^f}$ . Hence

$$1 - \frac{w_i^s}{w_i^f} \geq 1 - \frac{v'(n_{-i,i})}{w_{-i}^f} > \left( 1 - \frac{\lambda_{-i}}{\lambda_i} \right) \mu_{-i} \left( 1 - \frac{v'(n_{-i,i})}{w_{-i}^f} \right),$$

which means that  $\frac{dW}{dn_i^f} > 0$ . Therefore, it is never optimal to decrease  $n_i^f$  so much that type  $i$  works in the shadow economy.  $\square$



**Proof of Proposition 8.** First we will show how to obtain (16). The efficiency gain is straightforward. In order to obtain the redistribution gain, note that there are no distortions imposed on type  $-i$ , hence

$$\mu_{-i}\lambda_{-i} \left( U(c_{-i}^{SE}, n_{-i}^{SE}) - U(c_{-i}^M, n_{-i}^M) \right) = \mu_{-i}\lambda_{-i} (T_{-i}^M - T_{-i}^{SE}) = -\mu_i\lambda_{-i} (T_i^M - T_i^{SE}).$$

Summing up the terms results in (16). In order to derive thresholds, recall that  $H(w^s) = U(w^s g(w^s), g(w^s))$ . The efficiency gain is given by

$$\mu_i\lambda_i \left( H(w_i^s) - U\left(w_i^f n_i^M, n_i^M\right) \right),$$

it is strictly increasing in  $w_i^s$  and positive for  $w_i^s > \bar{w}_i^s$ . Note that by (51)  $n_i^M$  will always be distorted (downwards if  $i = l$ , upwards if  $i = h$ ). Hence,  $U(w_i^f n_i^M, n_i^M) < H(w_i^f)$  and the threshold  $\bar{w}_i^s$  is strictly lower than  $w_i^f$ .

Using the binding  $IC_{-i,i}$  constraint, we can express the redistribution gain as

$$\mu_i\mu_{-i}(\lambda_i - \lambda_{-i}) \left( U\left(w_i^f n_i^M, \frac{w_i^f}{w_{-i}^f} n_i^M\right) - H(w_{-i}^s) \right).$$

It is strictly decreasing in  $w_{-i}^s$  and is positive for  $w_{-i}^s < \bar{w}_{-i}^s$ . Since  $\frac{w_i^f}{w_{-i}^f} n_i^M \neq g(w_{-i}^f)$ , it is true that  $U(w_i^f n_i^M, \frac{w_i^f}{w_{-i}^f} n_i^M) < H(w_{-i}^f)$  and the threshold  $\bar{w}_{-i}^s$  is strictly lower than  $w_{-i}^f$ .  $\square$

## A Proofs from Section 3

**Proof of Lemma 1.** The single-crossing requires that  $\frac{d}{di} \left( \frac{\partial V_i(y^f, T)/\partial y^f}{\partial V_i(y^f, T)/\partial T} \right) < 0$ . Suppose that  $v'(\frac{y^f}{\phi_i}) < \psi_i$ . Then the agent supplies no informal labor and the indirect utility function  $V$  is just the utility function  $U$  evaluated at the formal allocation. Since  $v'$  is increasing, the single crossing holds in this case. When  $v'(\frac{y^f}{\phi_i}) \geq \psi_i$ , then the optimal provision of informal labor means that  $v'(n_i) = w_i^f$ , which implies  $\frac{\partial V_i(y^f, T)/\partial y^f}{\partial V_i(y^f, T)/\partial T} = \frac{w_i^s}{w_i^f}$ . Therefore the single crossing condition requires that  $\frac{d}{di} \left( \frac{w_i^s}{w_i^f} \right) < 0$ .  $\square$

**Proof of Proposition 9.** First note that the incentive compatibility requires that if  $\frac{d}{dj} V_i(y_j^f, T_j) \Big|_{j=i}$  exists, it is equal to 0. Otherwise type  $i$  can improve welfare by changing income marginally, so the allocation is not incentive compatible. Hence, if  $\frac{d}{di} V_i(y_i^f, T_i) = \frac{d}{dj} V_i(y_j^f, T_j) \Big|_{j=i} + \frac{d}{di} V_i(y_i^f, T_i) \Big|_{j=i}$  exists, it is equal to  $\frac{d}{di} V_i(y_i^f, T_i) \Big|_{j=i} = \left( \frac{w_i^f}{w_i^f} n_i^f + \frac{w_i^s}{w_i^s} n_i^s \right) v'(n_i)$ . We call this derivative a marginal information rent and denote it simply by  $\dot{V}_i$ .

By Milgrom and Segal (2002) (see their 10th footnote and Theorem 2), we can represent the utility schedule for any  $i < 1$  as an integral of marginal information rents

$$V_i(y_i^f, T_i) = V_0(y_0^f, T_0) + \int_0^i \dot{V}_j dj,$$

Moreover, the utility schedule  $V_i$  is continuous everywhere and differentiable almost everywhere.

Now we will show that the allocation is not incentive compatible if the formal income is decreasing in type. Suppose that the allocation is incentive-compatible and that there are two types  $a < b$  such that  $y_a^f > y_b^f$ . By the incentive compatibility, we have

$$V_a(y_a^f, T_a) \geq V_a(y_b^f, T_b). \quad (52)$$

$\frac{d}{di} V_i(y^f, T)$  is increasing in  $y^f$ . To see it, note that

$$\frac{d}{di} V_i(y^f, T) = \left( \rho_i^f \frac{y^f}{w_i^f} + \rho_i^s \max \left\{ g(w_i^s) - \frac{y^f}{w_i^f}, 0 \right\} \right) v'(n_i),$$

where  $g$  is the inverse function of  $v'$ . The single-crossing implies that  $\rho_i^f > \rho_i^s$ , so the right hand side is increasing in  $y^f$ .

Since  $y_a^f > y_b^f$ , for each type  $i$  it is true that  $\frac{d}{di} V_i(y_a^f, T_a) > \frac{d}{di} V_i(y_b^f, T_b)$ . It implies

$$V_b(y_a^f, T_a) - V_a(y_a^f, T_a) = \int_a^b \frac{d}{di} V_i(y_a^f, T_a) di > \int_a^b \frac{d}{di} V_i(y_b^f, T_b) di = V_b(y_b^f, T_b) - V_a(y_b^f, T_b). \quad (53)$$

Summing (52) and (53) results in

$$V_b(y_a^f, T_a) > V_b(y_b^f, T_b),$$

which contradicts the incentive-compatibility. Therefore, a nondecreasing formal income schedule is necessary for incentive compatibility. Conversely, suppose that the local incentive constraints hold and the formal income schedule is nondecreasing. Then for any two types  $a < b < 1$

$$V_b(y_b^f, T_b) - V_a(y_a^f, T_a) = \int_a^b \frac{d}{di} V_i(y_i^f, T_i) di \geq \int_a^b \frac{d}{di} V_i(y_a^f, T_a) di = V_b(y_a^f, T_a) - V_a(y_a^f, T_a), \quad (54)$$

which implies

$$V_b(y_b^f, T_b) \geq V_b(y_a^f, T_a).$$

We can use the same reasoning, but bound the utility difference on the left hand side of (54) from

above by  $\int_a^b \frac{d}{di} V_i(y_b^f, T_b) di$  to get

$$V_a(y_a^f, T_a) \geq V_a(y_b^f, T_b).$$

We cannot use this argument when  $b = 1$  and  $w_1^f$  is unbounded, but then by continuity of  $V_i$  we have  $\lim_{b \rightarrow 1} \{V_b(y_b^f, T_b) - V_b(y_a^f, T_a)\} \geq 0$ .  $\square$

**Proof of Theorem 1.** First we will derive  $D_i^f$  and  $D_i^s$  term. Then we will show that conditions from the theorem are necessary. Finally we will prove sufficiency.

Suppose that  $i \in \mathcal{F}$ . A perturbation of formal income changes the marginal information rent of type  $i$  by

$$\frac{\partial \dot{V}_i}{\partial y_i^f} = (1 - t_i) \rho_i^f \left(1 + \frac{1}{\zeta_i}\right). \quad (55)$$

where  $\zeta_i = \frac{v'(n_i)}{n_i v''(n_i)}$  is the elasticity of labor supply. This change of income affects the utility level of type  $i$  by  $\frac{dV_i}{dy_i^f} = 1 - \frac{v'(n_i)}{w_i^f}$ . By Proposition 9 the utility schedule has to be continuous, so we have to introduce additional change in tax  $T_i$  in order to keep the utility level of type  $i$  constant. We adjust the total tax paid by an agent of type  $i$  by  $dT_i = 1 - \frac{v'(n_i)}{w_i^f}$ . Note that  $dT_i$  is just equal the marginal tax rate  $t_i$ . This perturbation influences the tax revenue as if we were decreasing the formal income of type  $i$  while keeping the marginal tax rate fixed. Since we are interested in the tax revenue impact of the unit perturbation of the marginal information rent, we normalize  $dT_i$  by  $\frac{\partial \dot{V}_i}{\partial y}$ . In order to capture the tax revenue impact of perturbation of all agents of type  $i$ , we multiply this expression by  $\mu_i$  and get

$$D_i^f = dT_i \left( \frac{\partial \dot{V}_i}{\partial y_i^f} \right)^{-1} \mu_i = \frac{t_i}{1 - t_i} \left( \rho_i^f \left(1 + \frac{1}{\zeta_i}\right) \right)^{-1} \mu_i.$$

Suppose that  $i \in \mathcal{S}$ . Shadow labor is supplied according to  $v'(n_i) = w_i^s \implies n_i = g(w_i^s)$ . The marginal information rent can be expressed as

$$\dot{V}_i = \left( \frac{\dot{w}_i^f}{(w_i^f)^2} y_i^f + \frac{\dot{w}_i^s}{w_i^s} \left( g(w_i^s) - \frac{y_i^f}{w_i^f} \right) \right) w_i^s. \quad (56)$$

We marginally perturb  $y_i^f$ . The impact of the perturbation of the marginal information rent is

$$\frac{d\dot{V}_i}{dy_i^f} = (\rho_i^f - \rho_i^s) \frac{w_i^s}{w_i^f}.$$

As in the formal workers' case, the perturbation of  $y_i^f$  alone changes the utility level of type  $i$ . In order to keep the utility schedule continuous at  $i$ , we need to adjust the tax  $T_i$  such that the utility

of this type is unchanged. The required change of the tax is  $dT_i = 1 - \frac{v'(n_i)}{w_i^f}$ , which for the shadow worker equals  $\frac{w_i^f - w_i^s}{w_i^f}$ . By multiplying the tax revenue change with  $\mu_i$  and normalizing it with  $\frac{d\dot{V}_i}{dy^f}$ , we obtain the tax revenue cost of decreasing the marginal information rent of type  $i$ :

$$D_i^s = dT_i \left( \frac{\partial \dot{V}_i}{\partial y_i^f} \right)^{-1} \mu_i = \frac{w_i^f - w_i^s}{w_i^s} (\rho_i^f - \rho_i^s)^{-1} \mu_i.$$

If the interior formal income is nondecreasing, the interior allocation implied is incentive-compatible. The necessity of the conditions (30)-(33) was demonstrated in the main text. If these conditions do not hold, there exists a beneficial perturbation.

The conditions (30)-(33) are sufficient when they pin down the unique interior allocation. This happens when the cost of distortions is decreasing in the formal income of each type. Then the government's problem of choosing formal income of each type is concave. For formal workers it requires that  $\zeta_i$  is non-increasing in the labor supply, as then increasing the marginal tax rate  $t_i$  leads to an increase in  $D_i^f$ . For the marginal workers we need  $D_i^s > D_i^f$ , which is guaranteed by  $\frac{\rho_i^s}{\rho_i^f} > \zeta_i^{-1}$ . See the footnote 11 for the comment regarding the uniqueness of allocation for types for which  $D_i^s = \int_i^1 N_j dj$  holds.  $\square$

**Proof of Proposition 10.** We will examine the monotonicity of an interior formal income schedule separately for the formal, marginal and shadow workers.

The single-crossing condition implies that if the marginal tax rate is non-increasing in type, the formal income of workers in  $\mathcal{F}$  is increasing. By (30) the marginal tax rate satisfies

$$\frac{t_i}{1 - t_i} = \rho_i^f \left( 1 + \frac{1}{\zeta_i} \right) \frac{1 - M_i}{\mu_i} \mathbb{E}(1 - \omega_j | j > i). \quad (57)$$

Assumption 3(i) means that  $\mathbb{E}(1 - \omega_j | j > i) = \mathbb{E}\left(1 - \frac{\lambda_j}{\eta} \mid j > i\right)$  is non-increasing in  $i$ . Assumptions 3(ii) and 3(iii) imply that the rest of the right hand side of (57) is non-increasing in  $i$ . Hence,  $t_i$  is non-increasing and the interior formal income schedule is increasing in  $\mathcal{F}$ .

For any marginal worker  $i$  the formal income is fixed at  $w_i^f g(w_i^s)$ . The derivative of formal income with respect to type is

$$\dot{y}_i^f = \frac{dw_i^f g(w_i^s)}{di} = \dot{w}_i^f g(w_i^s) + w_i^f \dot{w}_i^s g'(w_i^s) = w_i^f g(w_i^s) \left( \rho_i^f + \rho_i^s \frac{w_i^s g'(w_i^s)}{g(w_i^s)} \right).$$

Notice that  $\frac{w_i^s g'(w_i^s)}{g(w_i^s)} = \frac{v'(n_i)}{n_i v''(n_i)} = \zeta_i$ . Hence, for any marginal worker  $\dot{y}_i^f \geq 0$  if and only if  $\rho_i^f + \rho_i^s \zeta_i \geq 0 \Leftrightarrow \frac{\rho_i^s}{\rho_i^f} \geq -\zeta_i^{-1}$ .

In the interior allocation all shadow workers have zero formal income. Hence, the formal income schedule is non-decreasing only if shadow workers are present exclusively at the bottom of the type

space. According to (32), a worker  $i$  belongs to  $\mathcal{S}$  in an interior allocation if and only if

$$\frac{w_i^f - w_i^s}{w_i^s} \leq \rho_i^f \frac{1 - M_i}{\mu_i} \left( 1 - \frac{\rho_i^s}{\rho_i^f} \right) \mathbb{E}(1 - \omega_j | j > i).$$

The left hand side is increasing in  $i$  by the single-crossing assumption. The right hand side is non-increasing by assumptions 3(i), 3(ii) and 3(iv).  $\square$

**Proof of Proposition 2.** We will show that under the assumptions made the interior allocation is such that bottom types do not work in the shadow economy, while some types above them do. This leads to the income schedule locally decreasing in type.

Let's compute the term  $\int_i^1 N_j dj$ . By (33) we know that  $\eta = \mathbb{E}\{\lambda_i\} = 1$ . Hence  $\int_i^1 N_j dj = \int_i^1 (1 - \lambda_j) dj$  and the derivative of this term is  $\frac{\partial \int_i^1 N_j dj}{\partial i} = \lambda_i - 1$ .

The term  $D_i^s$  is

$$D_i^s = \left( \frac{w_0^f}{w_0^s} e^{(\rho^f - \rho^s)i} - 1 \right) (\rho^f - \rho^s)^{-1}.$$

By (32) any type  $i$  is a shadow worker in the interior allocation if and only if  $\int_i^1 N_j dj \geq D_i^s$ . We can rewrite this inequality as

$$w_0^s \geq \frac{e^{(\rho^f - \rho^s)i}}{1 + (\rho^f - \rho^s) \int_i^1 N_j dj} w_0^f.$$

Denote the right hand side by  $X_i$ . Note that  $X_0 = 1$ , which together with  $w_0^f > w_0^s$  implies that the bottom types do not work in the shadow economy and by the Assumption 1(iii) have a positive formal income.

We define the threshold  $\bar{w}_0^s$  as  $\min_{i \in [0,1]} X_i$ . In order to see that  $\bar{w}_0^s < w_0^f$ , let's compute the derivative of  $X_i$  :

$$\dot{X}_i = (\rho^f - \rho^s) e^{(\rho^f - \rho^s)i} \left( 2 - \lambda_i + (\rho^f - \rho^s) \int_i^1 N_j dj \right).$$

Note that  $\dot{X}_0 = (\rho^f - \rho^s) (2 - \lambda_0) < 0$ , so  $X_i$  is decreasing at the bottom type and  $\min_{i \in [0,1]} X_i < X_0 = w_0^f$ . Therefore, whenever  $w_0^f > w_0^s > \bar{w}_0^s$ , the bottom types have a positive formal income, while some types above them work in the shadow economy and have no formal income.  $\square$

## Proof of Theorem 2.

*Proof.* There are three cases we should consider, depending on whether the interior formal income is increasing, locally decreasing, or increasing but not strictly. In the first case (strictly increasing

schedule) by Theorem 1 the interior allocation is optimal. In the second case (locally decreasing schedule) by Theorem 2 we need to use the optimal bunching condition (35). Below we derive this condition formally. In the third case the interior income schedule is non-decreasing with flat parts. By Theorem 1 the interior allocation is optimal. However, Theorem 2 says that the flat parts of the income schedule should be consistent with the optimal bunching condition (35). We will show that those two approaches are equivalent.

Suppose that the formal income schedule  $y^f$  is constant at the segment of types  $[a, b]$ . Let's marginally decrease the formal income of types  $[a, b]$ . Since we don't change the allocation of types below  $a$ , we have to make sure that  $V_a$  is unchanged - otherwise the utility schedule becomes discontinuous. Together with the cut of the formal income, we have to introduce a change in the total tax paid at this income level  $dT_a = 1 - \frac{v'(n_i)}{w_i^f} = t_{a-}$ . Since all types  $[a, b]$  are affected, the tax revenue loss is equal to

$$t_a(M_b - M_a). \quad (58)$$

Although this perturbation does not affect the utility of type  $a$ , it does affect the utility of all other bunched types. The utility impact of the perturbation of some type  $i \in (a, b)$  equals  $dU_i = 1 - \frac{v'(n_i)}{w_i^f} - dT_a = \frac{v'(n_a)}{\phi_a} - \frac{v'(n_i)}{\phi_i}$ . The welfare loss of bunched agents due to this utility change is

$$\int_a^b \Delta MRS_i \omega_i d\mu, \quad (59)$$

where  $\Delta MRS_i = \frac{v'(n_a)}{\phi_a} - \frac{v'(n_i)}{\phi_i}$ . Having the fiscal and welfare loss at the kink, we can add them into a cost of increasing distortions at the bunch  $[a, b]$ . We normalize the sum by  $t_{b+} - t_{a-}$ , which makes sure that the perturbation results in the unit change of the utility of type  $b$ , and we obtain (34). As the perturbation results in a uniform utility change of agents above the bunch, we can use the standard term (27) in order to obtain the optimal bunching condition (35).

Suppose that the interior formal income schedule is flat on the segment  $[a, b]$ . We will prove the equivalence of the interior optimality conditions and the optimal bunching condition. Let's consider the following sequence of perturbations. First, decrease the marginal information rent of agent  $a$  such that the formal income of this type falls by a unit. Take a marginally higher type and again perturb the marginal information rent such that the formal income of this agents is decreased by a unit as well. Do it until you reach type  $b$ . Note that incentive compatibility is preserved at each stage, since the formal income is always non-decreasing. The aggregate welfare impact of this sequence of perturbations is

$$W_{interior} = \int_a^b \frac{\partial \dot{V}_i}{\partial y} \left( D_i - \int_i^1 N_j dj \right) di,$$

where  $D_i \equiv \begin{cases} D_i^f & \text{if } i \in \mathcal{F} \\ D_i^s & \text{if } i \in \mathcal{S} \end{cases}$ . We do not need to consider the marginal workers, because their formal income is increasing (see the proof of Proposition 10), hence they cannot be bunched. We can decompose  $W_{interior}$  into three components

$$W_{interior} = \underbrace{\int_a^b \frac{\partial \dot{V}_i}{\partial y} D_i di}_{X_1} - \underbrace{\int_a^b \frac{\partial \dot{V}_i}{\partial y} \int_i^b N_j dj di}_{X_2} - \underbrace{\int_a^b \frac{\partial \dot{V}_i}{\partial y} \int_b^1 N_j dj di}_{X_3}.$$

Note that  $D_i = \frac{1-MRS_i}{\frac{\partial \dot{V}_i}{\partial y}} \mu_i$ , hence  $X_1 = \int_a^c (1 - MRS_i) \mu_i di$ . We observe that  $\frac{\partial \dot{V}_i}{\partial y} = \frac{\partial^2 V_i}{\partial i \partial y} = -MRS_i$  and we integrate  $X_2$  by parts

$$X_2 = - \int_a^b MRS_i \int_i^b N_j dj di = - \left( \left[ MRS_i \int_i^b N_j dj \right]_a^b + \int_a^b MRS_i N_i di \right) = - \int_a^b (MRS_i - MRS_a) N_i di.$$

We simply integrate  $X_3$

$$X_3 = - \int_a^b MRS_i \int_b^1 N_j dj di = - (MRS_b - MRS_a) \int_b^1 N_j dj.$$

Now by summing and rearranging the terms we get

$$\begin{aligned} W_{interior} &= X_1 - X_2 - X_3 \\ &= \int_a^b (1 - MRS_i) d\mu + \int_a^b (MRS_i - MRS_a) (1 - \omega_i) d\mu + (MRS_b - MRS_a) \int_b^1 N_j dj \\ &= \int_a^b (1 - MRS_a) d\mu + \int_a^b (MRS_a - MRS_i) \omega_i d\mu + (MRS_b - MRS_a) \int_b^1 N_j dj \\ &= t_{a-} (M_b - M_a) + \int_a^b \Delta MRS_i \omega_i d\mu + (t_{a-} - t_{b+}) \int_b^1 N_j dj = (t_{b+} - t_{a-}) \left( D_{a,b} - \int_b^1 N_j dj \right). \end{aligned}$$

Since  $t_{b+} - t_{a-} = \frac{v'(n_a)}{\phi_a} - \frac{v'(n_b)}{\phi_b} > 0$ , the sequence of interior optimality conditions is equivalent to the optimal bunching condition (35).  $\square$

**Proof of Proposition 11.** If the interior allocation is incentive-compatible, the claim holds. Suppose that it is not the case, i.e. there is a kink in the tax schedule. In this case incentive compatibility constrains the government from reducing the utility of agents above kink as much as in the interior case. Since  $G$  is concave, it means that  $N_j$  terms for  $j$  above the kink is weakly higher and the government's will to impose distortions does not decrease. If there are shadow workers at the bottom and the curve  $\int_i^1 N_j dj$  shifts upwards, then even more types will be bunched at zero formal income at the bottom.

Let's think about shadow workers which are not at the bottom of the type space. The continuity assumptions guarantee that  $D_i^s$  and  $\int_i^1 N_j dj$  terms are continuous in type. It implies that before any set of shadow workers that are not at the bottom of the type space is a marginal worker.

Consider an interior allocation with a bunch of shadow workers at some positive formal income level. If we flatten the interior formal income schedule in order to make it non-decreasing (as in Figure 6), the first type in the bunch (type  $\bar{a}$ ) will be a marginal worker ( $\frac{v'(\frac{y_{\bar{a}}}{w_{\bar{a}}^f}}{w_{\bar{a}}^s} = 1$ ), while all the other types with this level of formal income will be shadow workers ( $\frac{v'(\frac{y_{\bar{a}}}{w_i^f}}{w_i^s} < 1, i > \bar{a}$ ). To see this, note that  $\frac{\partial}{\partial_i} \left( \frac{v'(\frac{y_{\bar{a}}}{w_i^f})}{w_i^s} \right)$  is negative by  $\frac{\rho_i^s}{\rho_s} > -\zeta^{-1}$ . So far we discussed what happens at the flattened income schedule. The optimal income schedule involves no less distortions, so the shadow workers will not cease to supply shadow labor.  $\square$

**Proof of Corollary 2.** It is just an interior optimality condition for the shadow worker (32). By Lemma 11, all the shadow workers from the interior allocation are shadow workers in the optimum.  $\square$

## B The estimation of the factor $F_i$ and top earners Pareto distribution.

Here we present the variables included in the vector  $X_i$  and the parameter estimates of  $\beta$  and  $\gamma$  obtained from the specification given by (44). Table 4 lists the variables included in  $X_i$  with its corresponding description and associated category. The parameter estimates are presented in Table 5. Finally, table 6 presents the estimate of the Pareto distribution for top earners.



Table 4: Variables included in  $X_i$ 

Variable	Description	Values
Worker characteristics		
Gender	Dummy variable equal to 1 for women	0-1
Age	Age of the worker	16-90
Age <sup>2</sup>	Age squared	
Ed years	Number of education years	0-26
Degree	Highest degree achieved	1-5
		1 - no degree
		5 - postgraduate degree
Y work	Number of months worked in the last year	1-12
Experience	Number of months worked in the last job	0-720
First job	Dummy for the first job (1 if it is the first job)	0-1
Production unit (firm) characteristics		
Sector Man	Dummy for the manufacturing sector	0-1
Sector Fin	Dummy for financial intermediation	0-1
Sector ret	Dummy for the sales and retailers sector	0-1
Big city	Dummy for a firm in one of the two largest cities	0-1
Size	Categories for the number of workers	1-9
		1 - One worker
		9 - More than 101 workers
Production unit (Type of job) characteristics		
Lib	Dummy for a liberal occupation	0-1
Admin	Dummy for an administrative task	0-1
Seller	Dummy for sellers and related	0-1
Services	Dummy for a service task (bartender ..)	0-1
Worker-firm relationship		
Union	Dummy for labor union affiliation (1 if yes)	0-1
Agency	Dummy for agency hiring (1 if yes)	0-1
Seniority	Number of months of the worker in the firm	0-720

Table 5: Estimation results

Parameter	Point estimate	std. error	t-statistic	95% conf. interval	
$\gamma_0^f$	6.859	0.033	211.9	6.89	7.02
$\gamma_0^s - \gamma_0^f$	0.102	0.032	-3.2	-0.16	-0.04
$\gamma_1^s$	0.682	0.037	12.6	0.648	0.716
$\beta$ -Gender	-0.077	0.005	-11.6	-0.06	-0.04
$\beta$ -Age	0.025	0.001	13.1	0.01	0.02
$\beta$ -Age <sup>2</sup>	0.000	0.000	-8.8	0.00	0.00
$\beta$ -Ed years	0.037	0.002	15.4	0.02	0.03
$\beta$ -Degree	0.156	0.005	21.1	0.10	0.12
$\beta$ -Sector Man	-0.098	0.006	-11.9	-0.08	-0.06
$\beta$ -Sector Fin	0.156	0.015	6.9	0.08	0.14
$\beta$ -Sector ret	-0.150	0.006	-16.9	-0.11	-0.09
$\beta$ -Big city	0.010	0.007	1.0	-0.01	0.02
$\beta$ -Size	0.032	0.001	18.7	0.02	0.02
$\beta$ -Union	0.126	0.010	8.3	0.07	0.11
$\beta$ -Agency	-0.144	0.005	-18.3	-0.11	-0.09
$\beta$ -Seniority	0.001	0.000	17.9	0.00	0.00
$\beta$ -Y work	0.029	0.001	18.4	0.02	0.02
$\beta$ -First job	-0.053	0.008	-4.7	-0.05	-0.02
$\beta$ -Experience	0.000	0.000	5.3	0.00	0.00
$\beta$ -Lib	0.074	0.013	3.9	0.03	0.08
$\beta$ -Admin	-0.272	0.009	-19.9	-0.20	-0.17
$\beta$ -Seller	-0.186	0.014	-9.2	-0.15	-0.10
$\beta$ -Services	-0.267	0.009	-19.3	-0.20	-0.16

Table 6: Pareto distribution estimates

Parameter	Point estimate	std. error	z-statistic	95% conf. interval	
Shape parameter	1.81	0.0018	953.34	1.806	1.833