Carnegie Mellon
Tepper
SCHOOL OF BUSINESS

DISSERTATION

Submitted in partial fulfillment of the requirements
for the degree of

DOCTOR OF PHILOSOPHY
ECONOMICS

Titled
"ESSAYS ON MACROECONOMICS AND PUBLIC FINANCE"

Presented by
Antonio Andrés Bellofatto

Accepted by

Laurence Ales	**4/30/15**
Co-Chair: Prof. Laurence Ales	**Date**
Christopher Sleet	**4/30/15**
Co-Chair: Prof. Christopher Sleet	**Date**

Approved by The Dean

Robert M. Dammon	**4/30/15**
Dean Robert M. Dammon	**Date**

Essays on Macroeconomics and Public Finance

by

Antonio Andrés Bellofatto

Submitted in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

at

Tepper School of Business
Carnegie Mellon University

May 6, 2015

Committee: Laurence Ales (co-chair), Christopher Sleet (co-chair), Stephen Spear and Sevin Yeltekin

External Reader: Christopher Telmer

A Maru, Cata, Mamá y Papá.

Abstract

This dissertation contains three chapters and focuses on the optimal design of fiscal policy, both from a theoretical and from a quantitative perspective.

In the first chapter, "**Wealth Taxation and Life Expectancy**," I address the optimal taxation of wealth in a class of dynastic overlapping-generations economies with heterogeneous mortality risk. Working individuals are indexed by skills which are private information. Skills not only determine earning abilities but also correlate with survival probability, so that more productive agents on average live longer. The analysis distinguishes between the tax treatment of two possible sources of wealth, namely, savings and bequests, and points to the mortality gradient as a crucial determinant for optimal wealth taxation. Specifically, due to differential mortality: (a) earned wealth should be marginally taxed, (b) transferred wealth via bequests should be marginally subsidized, and (c) marginal tax schedules on bequests and inter-vivos transfers should be separated. I calibrate the model to U.S. data and quantitatively evaluate its tax implications. For the median worker, mortality differences create a force for marginally taxing capital mortality differences create a force for marginally taxing savings by up to 1.7%, and for marginally subsidizing bequests by as much as 3.4%. These figures are robust to the value of the societal intergenerational discount factor and can yield significant welfare gains.

In the second chapter, "**Taxing Atlas: Using Firm Data to Derive Optimal Income Tax Rates**" (joint with Laurence Ales and Jessie J. Wang), we analyze the optimal taxation of top labor incomes. Top income earners are modeled as managers who are heterogeneous across skills and operate a span-of-control technology, as in Rosen (1982). Managers privately observe their skill level, which increases the productivity of both effort and supervision, thus creating a scale-of-operations effect. We characterize optimal taxes in this environment and identify novel determinants linked to firm technology. Our main result is that to be consistent with U.S. firm data, the optimal top income tax rate should be roughly in line with the U.S. tax code, in contrast to previous results in the literature.

In the third chapter, "**Regional State Capacity and the Optimal Degree of Fiscal Decentralization**" (joint with Martín Besfamille), we study the optimal degree of fiscal decentralization in a federation. In our environment, regional governments are characterized by two dimensions of state capacity; namely, administrative and fiscal. These gauge the ability to deliver public goods and to raise tax revenues, respectively. Two regimes are compared: partial and full decentralization. Under partial decentralization, regional governments have no tax powers and

rely on central bailouts to refinance incomplete projects. Under full decentralization, regional governments refinance incomplete projects through capital taxes, in a context of tax competition. We show how the optimal degree of fiscal decentralization hinges on the relative magnitudes of each type of capacity. Specifically, for sufficiently low levels of fiscal capacity, bailing out regional governments is optimal regardless of the level of administrative ability. However, a combination of low levels of administrative capacity and high levels of fiscal capacity calls for fully decentralizing tax powers.

Acknowledgements

I am indebted to my advisers Laurence Ales, Chris Sleet, Steve Spear, and Sevin Yeltekin for priceless advice and constant support throughout this project. I would also like to thank Martín Besfamille, María Marta Ferreyra, Yaroslav Kryukov, Finn Kydland, Nicolas Petrosky-Nadeau, Jack Stecher, Chris Telmer, Shu Lin Wee, and Ariel Zetlin-Jones for valuable comments. Marcelo Fernández, Fabrizio Orrego, Carlos Ramírez, María Recalde, and Ben Tengelsen provided useful suggestions on earlier drafts. I also acknowledge financial support from the William Larimer Mellon Fellowship at Carnegie Mellon University.

Contents

Chapter 1

Wealth Taxation and Life Expectancy

1.1 INTRODUCTION

This paper is motivated by two observations regarding the interaction between life expectancy, wealth and the tax code. First, individuals' life spans affect the composition of their own wealth and that of their heirs: all else equal, workers with longer life expectancy tend to save more to smooth consumption, draw down their wealth more slowly during retirement, and bequeath later.[1] Accordingly, the impact of wealth taxation, in the form of savings and inheritance taxes, can vary substantially across individuals facing different mortality risk. The second observation is the well documented fact that socioeconomic status positively correlates with life expectancy.[2] In turn, lifetime savings patterns and the timing of intergenerational transfers might be useful signals of earning abilities and, hence, of optimal tax liability.[3] How would an *optimal* tax system incorporate these well established facts? In particular, how should the different *sources* of wealth be taxed to account for the distinct savings patterns among the long-lived and the short-lived? What is the optimal way to exploit the socioeconomic mortality gradient in order to achieve redistributive objectives? Are these effects quantitatively relevant?

The goal of this paper is to provide guidance on these questions, by focusing on the optimal design of wealth taxes in environments with uncertain life spans and heterogeneous mortality risk. The main contributions are twofold. First, the paper distinguishes between the optimal tax treatment of the three possible sources of household wealth; namely, taxes on *earned* wealth (henceforth *capital* taxes), and taxes on *transferred* wealth via *bequest* and *inter-vivos* gifts. The analysis points to the socioeconomic mortality gradient as a crucial determinant, primarily shaping capital and bequest taxes. Second, I quantify the key forces behind optimal wealth taxes using U.S. data. I find that the magnitude of the effect of the mortality gradient on optimal wealth taxes is commensurate with the actual level of capital taxes among developed countries.

[1] See, e.g., De Nardi et al. (2006), De Nardi et al. (2009) and Piketty (2014), Chapter 11.

[2] A seminal contribution can be found in Kitagawa and Hauser (1973). Contemporaneous studies, such as Singh and Siahpush (2006), Waldron (2007) and Pijoan-Mas and Ríos-Rull (2014), indicate that mortality differentials across socioeconomic groups have substantially widened in recent decades.

[3] Following Akerlof (1978), such variables can be thought of as "tags."

I consider dynastic environments in which agents live for at most two periods and are altruistic towards their descendants. Working individuals are indexed by *skills*, which are private information as in Mirrlees (1971). Crucially, skills not only determine the earning abilities of the agents (i.e. the ability to transform effort into effective effort), but also correlate with their survival probabilities, so that more productive individuals on average live longer. I adopt a utilitarian normative criterion which values children's welfare more than parents themselves do through altruism. In this sense, the *social* level of altruism is larger than the *private* one. This is a common specification in intergenerational models of social insurance (see Phelan (2006) and Farhi and Werning (2007, 2010)) and the idea is that, with altruistic parents, children should be "double counted" under any utilitarian welfare criterion which attaches separate weights to parents and children.[4] Optimal tax instruments are only restricted by informational frictions and, hence, are non-unique. Throughout the paper, I focus on a particular tax implementation of the constrained efficient allocation featuring nonlinear labor income taxes and linear wealth taxes whose rates depend on labor income histories, as in Kocherlakota (2005).

The analysis begins with a two-period dynastic economy which illustrates the main forces at work. The first period is populated by a continuum of parents. Parents work, consume and produce a single descendant, who is born at the beginning of the second period. After their children are born, parents are hit by a survival shock and only some of them survive to the second period. Survivors and children only consume.

I derive three key properties for optimal wealth taxes. First, optimal marginal taxes on capital are *positive*. Second, the optimal marginal bequest tax formula can be decomposed into two *negative* terms: one driven by differences between private and social altruism coefficients, and one driven by the socioeconomic mortality gradient. Hereafter, I refer to these terms as a *Pigouvian* and a *differential mortality* term, respectively. Third, in general, different marginal tax rates should apply to bequests and inter-vivos transfers. This normative prescription establishes a central departure from the current U.S. federal tax code where estate and inter-vivos gifts taxes are "unified," in the sense that the same statutory marginal rates and (lifetime) exemption levels apply to both types of intergenerational transfers.[5]

To grasp the intuition behind these results, it is useful to consider two benchmarks. A first benchmark is the case in which mortality risk is homogeneous across the population and social and private levels of altruism coincide. Here, optimal wealth taxes are zero and a version of the classical Atkinson and Stiglitz (1976) result holds; that is, commodity taxation is superfluous when the government has access to a nonlinear income tax schedule. A second benchmark corresponds to a setting in which social altruism is higher than the private one, but there is still no heterogeneity among life-expectancies. Under this scenario, capital taxes are again

[4] Hammond (1988) and Harsanyi (1995), on the other hand, advocate for excluding all "external" preferences from the social welfare function.

[5] The "unification" of bequests and inter-vivos tax schedules was incorporated into the U.S. federal tax system via the 1976 tax reform. Of course, some gaps remain. For instance, by 2014 annual gifts below $14,000 are tax-free. See Gale et al. (2001), Chapter 1.

zero while marginal bequests and inter-vivos taxes are negative and progressive. Essentially, when society cares more about future generations than parents themselves, intergenerational transfers generate a positive externality that makes it optimal to encourage such transfers via progressive subsidies. These *Pigouvian* forces behind optimal taxes on intergenerational transfers are extensively discussed in Farhi and Werning (2010). With respect to their work, a novelty of the present analysis is the potential distinction between the tax treatment of post-mortem and inter-vivos transfers.

In the general case, the mortality gradient affects optimal wealth taxes through two channels: positive marginal capital taxes and a negative differential mortality term on marginal bequest taxes. The intuition is as follows. The presence of mortality differences has effects similar to heterogeneous tastes over contingent consumption. In a nutshell, before the survival shock hits, high (low) ability parents prefer allocations featuring relatively high (low) levels of dynastic consumption in the event of survival. The policy maker can thus exploit this fact to motivate more productive types to exert effort. Specifically, the social planner distorts the contingent consumption bundles that high types would obtain if they didn't work at their full potential: it provides low income dynasties with "too little" consumption in the survival state and "too much" consumption in the death state. These distortions are implemented via positive marginal capital taxes and through the differential mortality term on marginal bequest taxes.

The next step in the analysis is to evaluate the quantitative relevance of the normative prescriptions discussed previously in a calibrated version of the model. The numerical simulations permit an exploration of dimensions of the model which are hard to characterize analytically.

To lay the groundwork for the quantitative analysis, the two-period version of the model is extended to a dynastic overlapping generations economy with heterogeneous mortality risk. The new environment addresses an important caveat of the two-period framework, since that model ignores the effect of wealth taxes on work incentives of the children.[6] The economy lasts for an infinite number of periods and it is populated by overlapping generations of agents who live for at most two periods. When young, agents work, consume and produce a descendant who is born in the following period. Old agents (those who survive to the second period) only consume. At the beginning of each period, young agents draw a productivity shock which is private information. At the end of each period, young agents draw a publicly observable survival shock which determines whether they live for an additional period. Both productivity and survival shocks are independently and identically distributed across dynasties and time. The economy features no physical capital accumulation, which is a recurrent assumption in general equilibrium models of social insurance following the original treatment of Atkeson and Lucas (1992).[7] To solve the model I extend the method developed by Farhi and Werning (2007) to an environment with

[6]The issue is particularly relevant when it comes to inheritance tax design, given that heirs expecting a bequest might have less incentives to work. This hypothesis, also known as "Carnegie's conjecture," is empirically confirmed by Holtz-Eakin et al. (1993).

[7]For examples in the context of infinite horizon Mirrleesian economies, see Albanesi and Sleet (2006) and Farhi and Werning (2006). As discussed in the body of the paper, the absence of capital affects the levels but not the shape of optimal wealth taxes.

uncertain life spans, heterogeneous mortality risk, and a continuum of skill types. I focus on the properties of optimal wealth taxes at a steady state, which is guaranteed to exist as long as the level of social altruism is greater than the private one. The analytical characterization of the model proves that the main normative prescriptions of the two-period economy extend to the infinite horizon environment at a steady state.

A crucial object for the calibration of the model is the probability of survival across skill types. While a number of previous studies have correlated survival probabilities to *observable* socioeconomic characteristics (such as income or education), to the best of my knowledge few have estimated this correlation with *unobservable* skills.[8] I calibrate this object to U.S. data by exploiting the relationship between mortality outcomes and permanent income in the Health and Retirement Study, a biennial panel survey of individuals over 50 years old. The methodology relies on techniques from the survival analysis literature. I find evidence of significant heterogeneity in life expectancies across skills. For example, individuals of 50 years old at the top 1% of the ability distribution are expected to live almost 8.5 years more than individuals at the bottom 1%.

Using this estimate, I quantify the effect of mortality heterogeneity on optimal wealth taxes. The major findings are three. First, using plausible values of the coefficient of social altruism I find that, for the median worker, mortality differences create a force for marginally taxing capital between 1.3% and 1.7% and marginally subsidizing bequests between 2.5% and 3.4% on average. Second, the effect of mortality differences on optimal wealth taxes is fairly irresponsive to the level of societal altruism and, hence, to the normative criterion employed. In particular, capital wedges and the differential mortality term on bequest wedges virtually do not change when the social level of altruism raises by 20%. Third, all else equal, certain distortions vary substantially with the current survival state of the dynasty. Expected bequest taxes, for example, can change by as much as 20 percentage points depending on whether old agents in the dynasty are alive or not. This finding suggests that partial tax reforms based on conditioning taxes on survival histories are possibly powerful.

Finally, I evaluate the potential welfare gains from tying optimal wealth taxes to mortality differentials. An ideal upper bound for such gains is one where non-zero wealth taxation is *only* justified on the grounds of mortality differentials. This would require shutting down Pigouvian forces by equalizing social and private levels of altruism, thus posing a serious technical problem in the infinite horizon version of the model: in that case, the well known "immiseration" result in dynamic contracting frameworks is obtained and a non-degenerate steady state fails to exist.[9] For this reason, I compute gains in the two-period version of the model. I find that welfare gains vary between 0.015% and 0.020% of aggregate consumption, for a reasonable range of values of risk aversion and labor supply elasticities. These welfare gains are small but not negligible. In fact, these estimates turn out to be much larger than in previous studies which quantify the gains from using optimal nonlinear capital taxes, also in two-period settings (see below).

[8] One exception is the contemporaneous work of Hosseini and Shourideh (2014).
[9] For details, refer to Atkeson and Lucas (1992) and Farhi and Werning (2007).

RELATED LITERATURE

This paper contributes to a recent literature on optimal nonlinear wealth taxation with private information initiated by Kocherlakota (2005) and Albanesi and Sleet (2006). Unlike my paper, this literature either fails to distinguish among the the source of wealth being taxed, thus focusing on the design of a broad based wealth tax schedule, or centers attention on bequest taxation. One exception can be found in Shourideh (2012) who studies the determinants of optimal wealth taxation in the presence of capital income risk and discriminates between capital income from controlled businesses, outside the business and bequests. Regarding bequest taxation, Farhi and Werning (2010) study optimal estate taxation in a dynastic framework that explicitly weights welfare of future generations. They conclude that optimal marginal estate taxes should be negative and progressive and show that this result is robust to a number of extensions. With respect to this work, my paper incorporates uncertain life spans and heterogeneous mortality risk. These are important ingredients for analyzing inheritance taxes, given that post-mortem transfers necessarily embed an accidental component. In addition, this paper provides a quantitative analysis of tax policy.

Farhi and Werning (2013a) analyze optimal estate taxation under altruism heterogeneity and find that optimal estate taxes can be *positive* depending on the redistributive objectives of the policy maker. A similar result is obtained by Piketty and Saez (2013) who evaluate optimal inheritance taxation using linear or two-bracket tax structures, in environments in which heterogeneity come from labor income and inheritance. In both of these frameworks, the positive bequest taxation result is underpinned by two features: a particular source of heterogeneity which is not earning abilities, and a special normative criteria that departs from the utilitarian metric. Unlike these works, agents in this paper are only indexed by earning abilities (which are perfectly correlated with survival probabilities) and the social welfare function is always utilitarian.

Given that heterogeneous mortality risk is isomorphic to heterogenous preferences, this paper is also related to Golosov et al. (2013) who quantitatively evaluate the case for tying nonlinear capital taxation to savings preferences in a two-period framework.[10] A relevant difference with their work is that while they find negligible welfare gains from using optimal capital taxes (in the order of 0.00002% of aggregate consumption), I obtain gains that can be up to three orders of magnitude larger.

Finally, this paper adds to the literature on the policy implications of differential mortality. My work is particularly related to the contemporaneous paper of Hosseini and Shourideh (2014), who analyze the effect of mortality differentials on optimal income tax design in a Mirrleesian framework. A notable difference with this paper, is that the authors focus on the optimal insurance arrangement of a single cohort in a context without bequest motives. Other papers in this area have been primarily devoted to studying whether the mortality gradient may negatively impact the progressivity of the social security benefit system (see, e.g., Garret (1995), Panis and Lillard

[10] Saez (2002) also justifies non-zero capital taxation based on heterogeneous discount rates across earning abilities.

(1995) and Brown (2000)).[11] To the best of my knowledge, the link between differential mortality and wealth taxation has so far been unexplored.

The remainder of the paper is organized as follows. Section 1.2 presents the two-period model and discusses the key determinants of wealth taxation arising from differences in mortality risk. Section 1.3 introduces the infinite horizon environment used for quantification. Section 1.4 discusses the calibration and Section 1.5 presents the quantitative results. Section 2.9 concludes. All proofs are contained in the Appendix.

1.2 TWO-PERIOD MODEL

In this section I characterize the optimal tax system in a two-period framework. Optimal taxes are recovered from the solution to a mechanism design problem where agents report their types to a social planner and receive allocations as a function of such reports. This is a standard practice in the dynamic public finance literature.[12]

1.2.1 ENVIRONMENT

Consider an economy that lasts for two periods indexed by $t = \{1, 2\}$ and which is populated by a continuum of dynasties. In the first period, the economy is populated by a unit measure of *young parents*, each identified as the head of her corresponding dynasty. Each young parent works, consumes and produces a single descendant or *child*. Children are born at the beginning of $t = 2$ and only consume. After their children are born, young parents are hit by a survival shock and only some of them survive to period $t = 2$. Survivors, henceforth *old parents*, only consume. Parents are altruistic to their children, but not vice versa.

Dynasties are indexed by their young parent's *skill*, θ, which is drawn at the beginning of $t = 1$ from a known distribution F with support $\Theta = \left[\underline{\theta}, \bar{\theta}\right]$. Skills have two roles. First, skills determine the ability with which young parents transform effort n_1 into effective effort y_1 according to the linear technology $y_1 = \theta \cdot n_1$. Second, skills affect the probability that the initial young survive to the next period. Specifically, I define the probability of survival to period $t = 2$ by $P : \Theta \to [0, 1]$ and make the following assumption:

Assumption 1.2.1. *P' exists and it is strictly positive.*

By Assumption 1.2.1 parents face heterogeneous longevity risk and more productive individuals on average live longer. This specification captures the well established fact that socioeconomic status (here indexed by θ) positively correlates with life expectancy. The empirical validity of Assumption 1.2.1 is confirmed in Section 1.4.

[11] In the words of Milton Friedman: "Persons in high income classes have a higher life expectancy, and so will tend to receive benefits for a longer period of time." (Friedman (1972), page 35).

[12] See, e.g., Golosov et al. (2007).

Preferences of the head of a dynasty with skill θ are represented by the expected utility function

$$U(\{c, y_1\}; \theta) = u(c_1^y) - h\left(\frac{y_1}{\theta}\right) + P(\theta)\delta\Big(u(c_2^o) + \beta u(c_2^y)\Big) + (1 - P(\theta))\delta\beta u(\tilde{c}_2^y),$$

where $c \equiv (c_1^y, c_2^o, c_2^y, \tilde{c}_2^y) \in \mathbb{R}_+^4$, c_1^y is the consumption of the young parent, c_2^o is the consumption of the old parent, and $\tilde{c}_2^y$ and c_2^y denote, respectively, the consumption of the child if the parent dies or survives the second period. $\beta > 0$ is the altruism coefficient, or *intergenerational* discount factor, while $\delta \in (0, 1)$ is the *intertemporal* discount factor. It is assumed that u', $-u''$, h' and h'' exist and are positive, $u'(0) = \infty$ and $u'(\infty) = h'(0) = 0$.

An allocation in this economy is defined as a mapping $\{c, y_1\}$ where $c : \Theta \to \mathbb{R}_+^4$ and $y_1 : \Theta \to [0, \bar{y}]$ with $0 < \bar{y} < \infty$; i.e. an allocation specifies (contingent) consumption for each member of the dynasty and effective labor as a function of the productivity of young parents.

By Assumption 1.2.1 it follows that for any allocation $\{c, y_1\}$

$$\frac{\partial}{\partial \theta}\left(\frac{\partial U(\{c, y_1\}; \theta)/\partial c_2^i}{\partial U(\{c, y_1\}; \theta)/\partial \tilde{c}_2^y}\right) > 0, \quad \text{for } i = o, y. \tag{1.1}$$

Verbally, high (low) types relatively prefer allocations where the dynasty consumes more in the survival (death) state. The logic is that, given their mortality types, dynasties enjoy allocations with higher consumption in their most likely state of nature.

I assume that goods can be transferred between periods using a linear savings technology with gross rate of return $R > 0$. Young parents have initial endowments of the consumption good in the amount W_1. An allocation is said to be *resource feasible* if it satisfies the sequential resource constraints

$$\int \Big[c_1^y(\theta) - y_1(\theta)\Big] dF(\theta) + K_2 \leq W_1, \tag{1.2}$$

and

$$\int \Big[P(\theta)\Big(c_2^o(\theta) + c_2^y(\theta)\Big) + (1 - P(\theta))\tilde{c}_2^y(\theta)\Big] dF(\theta) \leq RK_2, \tag{1.3}$$

where K_2 is aggregate capital.

Equations (1.2) and (1.3) reduce to the intertemporal resource constraint:

$$\int \left[c_1^y(\theta) - y_1(\theta) + \frac{1}{R}P(\theta)\Big(c_2^o(\theta) + c_2^y(\theta)\Big) + \frac{1}{R}(1 - P(\theta))\tilde{c}_2^y(\theta)\right] dF(\theta) \leq W_1. \tag{1.4}$$

Effective effort and consumption allocations are observable, while productivity realizations θ and effort n_1 are private information to the dynasties, as in Mirrlees (1971). For an allocation to be implementable, agents have to be induced to report their types truthfully. This requires

$$U(\{c(\theta), y_1(\theta)\}; \theta) \geq U(\{c(\theta'), y_1(\theta')\}; \theta), \quad \forall \theta, \theta', \tag{1.5}$$

where $U(\{c(\theta'), y_1(\theta')\}; \theta)$ denotes the expected utility of a dynasty of type θ who reports θ'. Any allocation that satisfies (2.5) is said to be *incentive compatible*.

Social welfare is evaluated according to the social welfare function

$$SWF = \int \Bigg[u(c_1^y(\theta)) - h\left(\frac{y_1(\theta)}{\theta}\right) + P(\theta)\delta\Big(u(c_2^o(\theta)) + \hat{\beta}u(c_2^y(\theta))\Big) + (1 - P(\theta))\delta\hat{\beta}u(\tilde{c}_2^y(\theta))\Bigg] dF(\theta), \quad (1.6)$$

where $\hat{\beta} \geq \beta$ denotes the level of *social* altruism.

Assuming $\hat{\beta} \geq \beta$ implies that society cares about descendants at least as much as parents do. This assumption implicitly reflects that parents and children are included as separate entities into the social welfare function, so that children are "double counted" given that parents are altruistic. In the special case that $\hat{\beta} = \beta$, social welfare is identified with that of the initial dynast.

Constrained efficient allocations maximize (1.6) subject to the intertemporal resource constraint (1.4) and the incentive compatibility constraints (2.5). This problem, however, is intractable due to the double infinity of incentive compatibility constraints embedded in (2.5). Hence, in what follows I work with a relaxed planning problem by applying a first-order approach, where the original set of incentive compatibility constraints (requiring that truth telling be a global maximum for each type) is replaced by local first order conditions which ensure that truth telling is a local maximum for each type.

To simplify notation, let

$$w_2(\theta) \equiv u(c_2^o(\theta)) + \beta u(c_2^y(\theta)) \quad \text{and} \quad \tilde{w}_2(\theta) \equiv \beta u(\tilde{c}_2(\theta)) \quad (1.7)$$

denote the continuation utilities contingent on survival and death of the parent, respectively.

The relaxed version of the planning problem is

$$\max_{\{c, y_1\}} \; SWF \quad (1.8)$$

subject to

$$\mathcal{V}(\theta) = u(c_1^y(\theta)) - h\left(\frac{y_1(\theta)}{\theta}\right) + P(\theta)\delta w_2(\theta) + (1 - P(\theta))\delta\tilde{w}_2(\theta), \quad \forall\theta, \quad (1.9)$$

$$\mathcal{V}'(\theta) = h'\left(\frac{y_1(\theta)}{\theta}\right)\frac{y_1(\theta)}{\theta^2} + \delta P'(\theta)\left(w_2(\theta) - \tilde{w}_2(\theta)\right), \quad \forall\theta, \quad (1.10)$$

and the intertemporal resource constraint (1.4), with w_2 and $\tilde{w}_2$ defined by (1.7).

The following Lemma provides a set of sufficient conditions which guarantee that solutions to the relaxed planning problem are incentive compatible, and hence are also solutions to the original problem where (1.9) and (1.10) are replaced by (2.5).

Lemma 1.2.1. *Let $\{c, y_1\}$ be a solution to the relaxed planning problem (1.8), and suppose that*

$$\frac{dy_1(\theta)}{d\theta} \geq 0, \quad \frac{dc_1^y(\theta)}{d\theta} \geq 0, \quad \frac{dw_2(\theta)}{d\theta} \geq 0, \quad \textit{and} \quad \frac{d\tilde{w}_2(\theta)}{d\theta} \leq 0, \tag{1.11}$$

where w_2 and $\tilde{w}_2$ are defined by (1.7). Then $\{c, y_1\}$ is incentive compatible, i.e. satisfies (2.5).

Proof. See Appendix 1.7.1. □

As is standard in the literature, the monotonicity conditions in (1.11) are assumed to hold and are verified ex-post (by means of numerical simulations).[13]

1.2.2 WEDGES

At any given allocation, it is possible to define three distortions: a *capital* wedge, τ^k, a *bequest* wedge, τ^b, and an *inter-vivos* transfers wedge, τ^i. These are given by

$$(1 - \tau^k(\theta))\delta R u'(c_2^o(\theta)) \equiv u'(c_1^y(\theta)), \tag{1.12}$$

$$(1 - \tau^b(\theta))\delta R \beta u'(\tilde{c}_2^y(\theta)) \equiv u'(c_1^y(\theta)), \tag{1.13}$$

and

$$(1 - \tau^i(\theta))\beta u'(c_2^y(\theta)) \equiv u'(c_2^o(\theta)). \tag{1.14}$$

Capital, bequest, and inter-vivos wedges can be thought of as implicit marginal wealth taxes on earned wealth through savings, transferred wealth at death, and transferred wealth between the living, respectively. Hence these wedges distort the three possible sources of wealth accumulation.[14]

While the focus of this paper is on τ^k, τ^b, and τ^i, one can also define the labor wedge, τ^n, as

$$1 - \tau^n(\theta) \equiv \frac{h'\left(\frac{y_1(\theta)}{\theta}\right)}{\theta u'(c_1^y(\theta))}. \tag{1.15}$$

Nonzero distortions are caused either by binding incentive compatibility constraints or by differences between societal and private intergenerational discount factors.

1.2.3 TAX IMPLEMENTATION

In this section I describe a tax implementation of the social optimum. This establishes the connection between the wedges defined in the previous section and the properties of an optimal tax system.

[13] Recent examples of this approach in the new dynamic public finance literature include Farhi and Werning (2013b), Golosov et al. (2013) or Kapička (2013a).

[14] Clearly, there are other ways of expressing wealth distortions. The standard alternative in the dynamic public finance literature is to define an "ex-ante" savings distortion which does not depend on future states of nature. See, e.g., Golosov et al. (2007).

The implementation features a nonlinear labor income tax schedule and linear wealth taxes whose rates depend on the head of the dynasty's level of effective labor, as in Kocherlakota (2005). Specifically, wealth tax functions are given by $t_2^w : [0, \bar{y}] \to \mathbb{R}$ for $w = \{k, b, i\}$, where t_2^k is a linear tax of personal wealth, and t_2^b and t_2^i are linear bequests and inter-vivos taxes, respectively. Under this decentralization a dynasty of type θ solves the following problem:

$$\max_{\{c, y_1, k_2, g_2\}} \quad u(c_1^y) - h\left(\frac{y_1}{\theta}\right) + P(\theta)\delta\Big(u(c_2^o) + \beta u(c_2^y)\Big) + (1 - P(\theta))\delta\beta u(\tilde{c}_2^y) \tag{1.16}$$

subject to

$$c_1^y + k_2 = y_1 - T_1(y_1) + W_1, \tag{1.17}$$
$$c_2^o + g_2 = Rk_2(1 - t_2^k(y_1)) + SS_2(y_1), \tag{1.18}$$
$$c_2^y = g_2(1 - t_2^i(y_1)), \tag{1.19}$$
$$\tilde{c}_2^y = Rk_2(1 - t_2^b(y_1)) + \tilde{SS}_2(y_1), \tag{1.20}$$

where $T_1 : [0, \bar{y}] \to \mathbb{R}$ is a nonlinear income tax schedule, $SS_2 : [0, \bar{y}] \to \mathbb{R}$ and $\tilde{SS}_2 : [0, \bar{y}] \to \mathbb{R}$ are government transfers received at $t = 2$ if the head of the dynasty survives or dies, respectively, $k_2 \geq 0$ is the level of savings, and $g_2 \geq 0$ denotes the level of inter-vivos gifts.[15,16]

I start by defining a competitive equilibrium with taxes in this economy.

Definition 1.2.1. *A competitive equilibrium with taxes is an allocation for consumption and effective effort* $\{c(\theta), y_1(\theta)\}_{\theta \in \Theta}$*, a sequence of capital and inter-vivos gifts* $\{k_2(\theta), g_2(\theta)\}_{\theta \in \Theta}$ *and a tax system* $\Phi \equiv \{T_1, t_2^k, t_2^b, t_2^i, SS_2, \tilde{S}S_2\}$ *such that*

1. *Taking* Φ *as given, for each* θ, $\{c(\theta), y_1(\theta), k_2(\theta), g_2(\theta)\}$ *solves the problem of the* θ*-dynasty* (1.16).

2. *The government's budget constraint is balanced in every period, i.e.*

$$\int T_1(y_1(\theta))dF(\theta) = 0,$$

and

$$\int \Big[P(\theta)\Big(Rk_2(\theta)t_2^k(y_1(\theta)) + g_2(\theta)t_2^i(y_1(\theta)) - SS_2(y_1(\theta))\Big) + (1 - P(\theta))\Big(Rk_2(\theta)t_2^b(y_1(\theta)) - \tilde{SS}_2(y_1(\theta))\Big)\Big]dF(\theta) = 0.$$

[15] Note that t_2^b can be interpreted as either an estate or an inheritance tax. In reality, there is a difference. An estate tax is based on the value of the property owned by the deceased regardless of the identity of the beneficiary. In contrast, an inheritance tax applies to heirs and might be a function of several characteristics of the recipient such as her relationship with the donor, her income, etc. In this two-period model, however, there is no such distinction given that there is a single beneficiary per dynasty and t_2^b only depends on the parent's level of effective labor. With the exception of the U.S. and the U.K., among developed countries inheritance taxes are much more common than estate taxes at the federal level. See Cremer and Pestieau (2006).

[16] Transfers SS_2 and $\tilde{SS}_2$ can be thought of as representing two features of a Social Security system, namely, the retirement and survivors benefit programmes. A similar interpretation can be found in Shourideh and Troshkin (2012) and Golosov et al. (2013).

3. *The intertemporal resource constraint* (1.4) *holds, so that the goods market clears.*

Note that the government has no revenue requirements. This assumption is without loss of generality.

I say that an allocation $\{c, y_1\}$ is *implemented* by the tax system $\Phi \equiv \{T_1, t_2^k, t_2^b, t_2^i, SS_2, \tilde{SS}_2\}$ if there is a sequence $\{k_2(\theta), g_2(\theta)\}_{\theta \in \Theta}$ such that $\{c, y_1, k_2, g_2\}$ and Φ is a competitive equilibrium with taxes. The next proposition provides the implementation result.

Proposition 1.2.1. *Consider an optimal allocation* $\{c^*, y_1^*\}$ *that solves* (1.8) *and satisfies* (1.11). *Then* $\{c^*, y_1^*\}$ *can be implemented by a tax system* $\{T_1^*, t_2^{k*}, t_2^{b*}, t_2^{i*}, SS_2^*, \tilde{SS}_2^*\}$ *such that*

$$t_2^{w*}(y_1^*(\theta)) = \tau^{w*}(\theta), \tag{1.21}$$

for $w = \{k, b, i\}$, *where for all* θ, $\{\tau^{w*}(\theta)\}_{w=k,b,i}$ *are the wedges defined in* (1.12)-(1.14) *evaluated at* $\{c^*(\theta), y_1^*(\theta)\}$.

Proof. See Appendix 1.7.1. □

The following section analyzes the properties of optimal marginal wealth taxes. Given the equivalence between marginal taxes and wedges in (1.21), for the rest of the analysis I center the discussion around wealth wedges, which simplifies notation. Moreover, I occasionally alternate between the use of "wedge" and "marginal tax" throughout the exposition.

1.2.4 A BENCHMARK: HOMOGENEOUS MORTALITY RISK

Before presenting optimal wedge formulas in this environment, I discuss the case where mortality risk is homogeneous across the population; that is, when Assumption 1.2.1 does not hold. This is a useful benchmark as it isolates the effect of uncertain life spans on the optimal tax system.

Proposition 1.2.2 characterizes optimal wealth wedges in this framework. The proof is omitted, given that this case is nested by Proposition 1.2.3 below. In what follows, $\{\tau^{w*}(\theta)\}_{w=k,b,i}$ denote the wedges defined in (1.12)-(1.14) evaluated at a given solution to the relaxed planning problem (1.8).

Proposition 1.2.2. *Suppose that for all* θ, $P(\theta) = \bar{P}$ *for some constant* $\bar{P} \in (0, 1)$. *Let* $\{c^*, y_1^*\}$ *denote a solution to the relaxed planning problem* (1.8). *Then for all* θ, *optimal wealth wedges satisfy*

$$\tau^{k*}(\theta) = 0, \qquad \textit{and} \qquad \tau^{b*}(\theta) = \tau^{i*}(\theta) = -\left(\frac{\hat{\beta}}{\beta} - 1\right) \frac{u'(c_1^{y*}(\theta))}{\lambda} \leq 0,$$

where λ *is the Lagrange multiplier on the intertemporal resource constraint* (1.4).

Proposition 1.2.2 immediately implies that when mortality risk is homogeneous across the population and $\hat{\beta} = \beta$, optimal wealth wedges are zero. In this case, a version of the classical

Atkinson and Stiglitz (1976) result holds and wealth taxes are superfluous: the optimum can be implemented by relying only on a nonlinear income tax schedule.

A gap between $\hat{\beta}$ and β has two effects of τ^{b*} and τ^{i*}. First, it creates a force for marginally subsidizing both post-mortem and inter-vivos transfers to descendants. The idea is that when $\hat{\beta} > \beta$ society cares more about descendants than parents themselves, which makes it optimal to encourage intergenerational transfers. Second, any difference between intergenerational discount factors renders tax schedules on bequests and inter-vivos transfers progressive, in the sense that marginal taxes are increasing in θ. This property immediately follows from the concavity of the utility functions and the fact that consumption allocations are increasing in θ under (1.11).

Both of these effects are discussed in detail in Farhi and Werning (2010). The main departure from their model is that the authors assume that agents live for one period. Accordingly, in their framework, inter-vivos transfers are precluded by construction and bequests are purely intentional. An environment with uncertain life spans, on the other hand, not only allows for intergenerational transfers between the living, but also incorporates an essential difference between inter-vivos and post-mortem transfers; namely, that bequests are partly accidental while inter-vivos transfers are purely intentional.

However, absent variations in mortality risk, the basic results of Farhi and Werning (2010) apply to both bequests and inter-vivos transfers under this richer framework. Essentially, in the absence of differential mortality, bequest and inter-vivos gifts should be taxed "uniformly," in the sense that either type of transfer should be subject to the same progressive marginal tax schedules. This normative prescription has its counterpart in the U.S. federal tax code, where the same marginal rates and (lifetime) exemption levels apply to either bequests and inter-vivos transfers.

The intuition for merging inter-vivos and bequests taxes into a single schedule is the following. Under homogeneous mortality, the planner subsidizes transfers to younger generations only as long as the act of giving generates a positive externality. But the magnitude of the externality is related to the difference between $\hat{\beta} - \beta$ and not to the state of nature (death or survival) in which transfers take place. Hence, when mortality risk is uniform across dynasties, the same subsidies should apply for wealth passed on at death or while alive. Put differently, separating bequests from inter-vivos gifts taxes neither serves a corrective role, nor facilitates incentive provision. In the next subsection I show that this result critically hinges on the assumption that mortality risk is homogeneous across the population.

1.2.5 HETEROGENEOUS MORTALITY RISK

The following Proposition characterizes optimal wealth distortions in the presence of a positive mortality gradient.

Proposition 1.2.3. *Suppose* $\{c^*, y_1^*\}$ *solves the relaxed planning problem and satisfies* (1.11). *Then for all*

θ, *optimal wealth wedges satisfy*

$$\tau^{k*}(\theta) = \frac{\mu(\theta)}{f(\theta)} \frac{P'(\theta)}{P(\theta)} \frac{u'(c_1^{y*}(\theta))}{\lambda}, \tag{1.22}$$

$$\tau^{b*}(\theta) = -\left(\frac{\hat{\beta}}{\beta} - 1\right) \frac{u'(c_1^{y*}(\theta))}{\lambda} - \frac{\mu(\theta)}{f(\theta)} \frac{P'(\theta)}{1 - P(\theta)} \frac{u'(c_1^{y*}(\theta))}{\lambda}, \tag{1.23}$$

$$\tau^{i*}(\theta) = -\left(\frac{\hat{\beta}}{\beta} - 1\right) \frac{\delta R u'(c_2^{o*}(\theta))}{\lambda}, \tag{1.24}$$

where $\lambda > 0$ *is the Lagrange multiplier on the intertemporal resource constraint* (1.4) *and* $\mu(\theta) \geq 0$ *is the costate associated with* (1.10) *satisfying* $\mu(\underline{\theta}) = \mu(\bar{\theta}) = 0$.

Proof. See Appendix 1.7.1. □

The wedge formulas in Proposition 1.2.3 reveal that differential mortality directly impacts optimal capital and bequest taxes. These two channels are evident by the presence of the costate $\mu(\theta)$ in the optimal capital tax formula (1.22) and in the second term on the formula for bequest taxes (1.23). In essence, these terms reveal that the social planner can relax incentive constraints through the effective use of wealth taxes.

Note that the *Pigouvian* forces identified previously, which emerge from $\hat{\beta} \geq \beta$, are still present in the case of heterogenous mortality risk. However, while inter-vivos taxes only respond to such forces, optimal marginal bequest taxes can now be decomposed into a *Pigouvian* term (reacting to $\hat{\beta} - \beta$) and a *differential mortality* term (the rightmost term on (1.23)). Accordingly, the shapes of inter-vivos and bequest taxes will not typically coincide. This implies that these tax schedules should no longer be unified, as they were in the absence of heterogenous mortality risk. In particular, note that when $\hat{\beta} = \beta$ optimal policy prescribes zero marginal taxes on inter-vivos gifts but negative marginal taxes on bequests.[17]

To better understand the impact of mortality gaps on optimal capital and bequest taxes, suppose that $\hat{\beta} = \beta$, so that wealth taxation is only justified on the grounds of mortality differentials and $\tau^{k*}(\theta) \geq 0 \geq \tau^{b*}(\theta)$. Intuitively, differential mortality makes more productive types relatively prefer allocations where dynastic consumption is high under parental survival (see equation (1.1)). The planner exploits this fact to motivate more productive types to exert effort. In particular, relative to the full information benchmark, the planner allocates "too little" consumption to the dynasties in the survival state and "too much" consumption in the death state. Such distortions are implemented by combining positive marginal capital taxes with negative marginal bequest taxes. Hence, potential deviators get relatively large after-tax returns in their less likely state of

[17] When $\hat{\beta} > \beta$, it is hard to determine the relative magnitudes of bequest and inter-vivos wedges analytically. In subsequent sections I analyze this issue quantitatively.

nature (death) and relatively low returns in the most likely state (survival), which relaxes incentive constraints.[18]

1.2.6 EXTENSIONS AND QUALIFICATIONS

Given that optimal tax tax instruments are only restricted by informational frictions, the implementation of the social optimum is not unique. The previous analysis focuses on a particular implementation where individuals can only save using a risk free asset, and are subject to wealth taxes on accumulated capital and bequests.

A natural extension is to consider a decentralization in which individuals can trade annuity contracts to hedge against lifespan uncertainty. In Appendix 1.7.4 I develop such an extension and show that the optimal wealth taxes derived in the body of the paper are robust to the inclusion of annuities.

One qualification when interpreting the results is that individuals in the model live for at most two periods. As a consequence, there are two possible interpretations for the intergenerational transfer which takes place in the second period. Namely, this transfer could either be an inter-vivos gift or a bequest, depending on whether wealth is passed on before or after death, respectively. In the body of the paper I adopt the former interpretation. Under the latter, there would be two types of bequests in the model: *"early" bequests* made at the end of the first period (when parents die prematurely), and *"late" bequests* made at the end of the second period. In this case, optimal bequest taxes would be such that bequests are marginally subsidized at a rate that decreases with the age of the donor.

1.3 INFINITE HORIZON

This section introduces the infinite horizon version of the model, laying the groundwork for the quantitative analysis of subsequent sections. Unlike in the two-period version of the model, in this environment children work in every period. Hence, this richer formulation incorporates the effect of wealth taxes on work incentives of younger generations.

1.3.1 ENVIRONMENT

Consider an overlapping generations economy that lasts for $T = \infty$ periods indexed by $t \in \mathbb{N}$. Agents face uncertain life spans and live for at most two periods. When *young*, agents work, consume and produce a single descendant who is born in the following period. *Old* agents, i.e. those who survive to the second period of their lives, only consume. A unit measure of *initial old*

[18] It is straightforward to show that the implicit average marginal tax that any downward deviator would face out-of-the-equilibrium-path is larger than if he reported truthfully. Specifically, define $\tau^E(\theta';\theta) \equiv P(\theta)\tau^{k*}(\theta') + (1 - P(\theta))\tau^{b*}(\theta')$ as the dynasty's *average wealth distortion* for any type θ reporting θ'. Using (1.22) and (1.23) it it follows that for $\theta' \leq \theta$ $\tau^E(\theta;\theta) = 0$ while $\tau^E(\theta';\theta) \geq 0$.

individuals is alive at $t = 1$. Individuals are altruistic towards their descendants, so that the model allows for a dynastic interpretation where each initial old corresponds to the head of a dynasty.

Agents are subject to two types of idiosyncratic shocks: productivity and survival shocks. At the beginning period t, young agents draw a productivity shock θ_t from a distribution F with support $\Theta = \left[\underline{\theta}, \bar{\theta}\right]$ and density f. Similarly, at the end of period t, young agents draw a survival shock $s_t \in \{0, 1\}$ with probability $\pi(s_t)$. For any agent born at $t-1$, $s_t = 1$ if such an agent survives to t and $s_t = 0$ otherwise. Both productivity and survival shocks are i.i.d. across dynasties and time. I denote t-histories of productivity and survival shocks by $\theta^t \equiv (\theta_1, \theta_2, ..., \theta_t) \in \Theta^t$ and $s^t \equiv (1, s_2, ..., s_t) \in \{0, 1\}^t$.[19] Productivity realizations are private information to the agents, but survival shocks are publicly observable.

Dynasties are identified by their initial discounted expected utility entitlements $w_1 \in \mathcal{W}_1$ which are distributed according to the distribution Ψ_1 with density ψ_1. Let $y_t \in [0, \bar{y}]$ denote the level of effective effort of the young at t, and let $c_t^y \geq 0$ and $c_t^o \geq 0$ denote consumption of the young and the old in period t, respectively. An allocation of consumption and effective effort is defined by the sequence $\{c, y\} \equiv \{c_t^y, c_t^o, y_t\}_{t=1}^{\infty}$ where[20]

$$c_t^y : \mathcal{W}_1 \times \Theta^t \times \{0,1\}^t \to \mathbb{R}_+^t, \quad c_t^o : \mathcal{W}_1 \times \Theta^t \times \{0,1\}^t \to \mathbb{R}_+^t, \quad y_t : \mathcal{W}_1 \times \Theta^t \times \{0,1\}^t \to [0, \bar{y}].$$

As usual, productivity shocks determine the ability of the young agents at time t to transform effort into effective effort using the linear technology $y_t(w_1, \theta^t, s^t) = \theta_t n_t(w_1, \theta^t, s^t)$, where n_t is effort at time t. In addition, productivity shocks impact the probability that young agents survive to the next period. Specifically, let $\pi(s^t|\theta^{t-1})$ denote the conditional probability of drawing survival history s^t given skill shock realization θ^{t-1}. I assume that only *own* productivity realizations affect the probability of survival of the agents, so that $\pi(s_t|\theta^{t-1}) = \pi(s_t|\theta_{t-1})$. Hence, I can write

$$\pi(s^t|\theta^{t-1}) = \pi(s_1)\pi(s_2|\theta_1)...\pi(s_t|\theta_{t-1}),$$

with

$$\pi(s_t|\theta_{t-1}) = \begin{cases} P(\theta_{t-1}), & s_t = 1, \\ 1 - P(\theta_{t-1}), & s_t = 0, \end{cases}$$

where $P : \Theta \to [0, 1]$ denotes the probability of survival as a function of the skill realization. Like in the two-period model, I assume that P' exists and it is strictly positive.

Preferences of a w_1-dynasty over the allocation $\{c, y\}$ can be represented by the expected utility

[19] Without loss of generality, I assume that $s_1 = 1$ so that the initial old are alive when the economy starts.

[20] Note that c_t^o is measurable with respect to θ^t. Hence, it is assumed that the old learn the realization of the current young's skill when making their choices.

function

$$U(\{c,y\};w_1) = \sum_{t=1}^{\infty}\sum_{s^t}\int_{\Theta^t}(\beta\delta)^{t-1}\pi(s^t|\theta^{t-1})\Big[\mathbb{I}(s_t=1)u(c_t^o(w_1,\theta^t,s^t)) + \beta\left(u(c_t^y(w_1,\theta^t,s^t)) - h\left(\frac{y_t(w_1,\theta^t,s^t)}{\theta_t}\right)\right)\Big]f^t(\theta^t)d\theta^t. \quad (1.25)$$

where $\mathbb{I}(s_t=1)$ is an indicator function which equals one when $s_t=1$, $f^t(\theta^t)\equiv f(\theta_1)f(\theta_2)...f(\theta_t)$ denotes the density of θ^t, $\beta>0$, $\delta\in(0,1)$ and $\beta\delta<1$. I assume that u', $-u''$, h' and h'' exist and are positive, $u'(0)=\infty$ and $u'(\infty)=h'(0)=0$.

An allocation $\{c,y\}$ is said to be *resource feasible* if for all $t\in\mathbb{N}$

$$\int_{\mathcal{W}_1\times\Theta^t}\sum_{s^t}\pi(s^t|\theta^{t-1})\Big[\mathbb{I}(s_t=1)c_t^o(w_1,\theta^t,s^t) + c_t^y(w_1,\theta^t,s^t) - y_t(w_1,\theta^t,s^t)\Big]f^t(\theta^t)\psi_1(w_1)d\theta^t dw_1 + H = 0, \quad (1.26)$$

where $H\geq 0$ is a constant level of exogenous government purchases.

Note that there is no physical capital in this economy. This is a common assumption among infinite horizon models of social insurance in general equilibrium, following Atkeson and Lucas (1992).[21] The absence of capital considerably simplifies the analysis but does not affect the shape of optimal wealth wedges.[22]

Given that that productivity shocks are privately observed, any optimal mechanism should induce agents to report skill realizations truthfully. Define a reporting strategy $\sigma\equiv\{\sigma^t\}_{t=0}^{\infty}$, where $\sigma^t:\Theta^t\to\Theta$. An allocation is then said to be *incentive compatible* if, for all $w_1\in\mathcal{W}_1,\theta^t\in\Theta^t,\sigma^t\in\Theta^t$

$$U(\{c,y\};w_1) \geq \sum_{t=1}^{\infty}\sum_{s^t}\int_{\Theta^t}(\beta\delta)^{t-1}\pi(s^t|\theta^{t-1})\Big[\mathbb{I}(s_t=1)u(c_t^o(w_1,\sigma^t,s^t)) + \beta\left(u(c_t^y(w_1,\sigma^t,s^t)) - h\left(\frac{y_t(w_1,\sigma^t,s^t)}{\theta_t}\right)\right)\Big]f^t(\theta^t)d\theta^t. \quad (1.27)$$

An allocation is said to be *feasible* if it satisfies (1.26), (1.27) and delivers utility w_1 to the dynasties with initial entitlement w_1, i.e.

$$U(\{c,y\};w_1) = w_1. \quad (1.28)$$

[21] In particular, Farhi and Werning (2006) also adopt this assumption in a dynastic Mirrlees model with one-period lived agents.

[22] On the other hand, the levels of optimal wedges are potentially affected by the absence of physical capital. See below.

The social planner ranks allocations according to the social welfare function

$$SWF = \int_{\mathcal{W}_1} \sum_{t=1}^{\infty} \sum_{s^t} \int_{\Theta^t} (\hat{\beta}\delta)^{t-1} \pi(s^t|\theta^{t-1}) \Big[\mathbb{I}(s_t = 1) u(c_t^o(w_1, \theta^t, s^t)) + \hat{\beta} \left(u(c_t^y(w_1, \theta^t, s^t)) - h\left(\frac{y_t(w_1, \theta^t, s^t)}{\theta_t} \right) \right) \Big] f^t(\theta^t) \psi_1(w_1) d\theta^t dw_1, \quad (1.29)$$

where $\hat{\beta} \in (\beta, \frac{1}{\delta})$ is the coefficient of social altruism.

Constrained efficient allocations solve the dynamic mechanism design problem:

$$\max_{\{c,y\}} \; SWF \quad (1.30)$$

subject to (1.26), (1.27) and (1.28).

1.3.2 A RELAXED PLANNING PROBLEM

As shown by Atkeson and Lucas (1992), the planning problem in (1.30) has a recursive structure. At any period t, the state variable of this problem corresponds to the cross-sectional distribution of utility entitlements Ψ_t, say. Consequently, the solution to the planning problem defines a mapping Ω and a law of motion $\Psi_{t+1} = \Omega(\Psi_t)$. A *steady state* in this environment is defined as a distribution of continuation utility entitlements Ψ^* satisfying $\Psi^* = \Omega(\Psi^*)$. The existence of a non-degenerate steady state distribution hinges on $\hat{\beta} > \beta$. When $\hat{\beta} = \beta$, on the other hand, long-run inequality is unbounded and the classical "immiseration" result of dynamic contracting frameworks holds.[23]

Working with the distribution of utilities as a state variable poses some obvious challenges, as Ψ_t is an infinite dimensional object. In what follows I take an alternative route by focusing on a relaxed version of the original planning problem. Crucially, the solution to the relaxed problem coincides with the original one *at a steady state*, but the former admits a recursive formulation using a one-dimensional state variable. This approach essentially extends the method developed by Farhi and Werning (2007) to an environment with uncertain life spans, heterogeneous mortality risk and a continuum of skill types.

The relaxed version of the planning problem is obtained by replacing the original sequence of resource constraints in (1.26) by a single intertemporal resource constraint:

[23] While I do not provide a proof showing that a steady state exists, my numerical simulations indicate that this is the case (see Appendix 1.7.3). Farhi and Werning (2007) formally prove that the existence of a steady state is guaranteed under $\hat{\beta} > \beta$ in a dynastic environment with one-period lived agents.

$$\sum_{t=1}^{\infty} q^{t-1} \int_{\mathcal{W}_1 \times \Theta^t} \sum_{s^t} \pi(s^t|\theta^{t-1}) \Big[\mathbb{I}(s_t = 1) c_t^o(w_1, \theta^t, s^t) + c_t^y(w_1, \theta^t, s^t) - y_t(w_1, \theta^t, s^t) \Big] f^t(\theta^t) \psi_1(w_1) d\theta^t dw_1 = -H \sum_{t=1}^{\infty} q^{t-1}, \quad (1.31)$$

for some intertemporal price $q > 0$.

The original set of resource constraints implies (1.26), while the converse is true at a steady state. The relaxed planning problem can be written as

$$\max_{\{c,y,\hat{\lambda}\}} \int_{\mathcal{W}_1} \mathcal{L}(w_1) \psi(w_1) dw_1, \quad (1.32)$$

subject to (1.27) and (1.28), where

$$\begin{aligned} \mathcal{L}(w_1) \equiv \quad & \sum_{t=1}^{\infty} \sum_{s^t} \int_{\Theta^t} \pi(s^t|\theta^{t-1}) \left(\hat{\beta}\delta\right)^{t-1} \Big\{ \mathbb{I}(s_t = 1) u(c_t^o(w_1, \theta^t, s^t)) \\ & + \hat{\beta} \Big(u(c_t^y(w_1, \theta^t, s^t)) - h\left(\frac{y_t(w_1, \theta^t, s^t)}{\theta_t} \right) \Big) \\ & - \hat{\lambda} \left(\frac{q}{\hat{\beta}\delta} \right)^{t-1} \left[\mathbb{I}(s_t = 1) c_t^o(w_1, \theta^t, s^t) + c_t^y(w_1, \theta^t, s^t) - y_t(w_1, \theta^t, s^t) \right] \Big\} f^t(\theta^t) d\theta^t. \end{aligned} \quad (1.33)$$

where $\hat{\lambda} > 0$ is the multiplier on the intertemporal resource constraint (1.31).

Given w_1 and $\hat{\lambda}$, define a *component planning problem* by

$$\max_{\{c,y\}} \ \mathcal{L}(w_1) \quad (1.34)$$

subject to (1.27) and (1.28).

The advantage of working with the relaxed planning problem is that it allows for a simple recursive formulation at steady states, along the lines of Spear and Srivastava (1987). Specifically, first note that a steady state requires $q = \hat{\beta}\delta$; otherwise aggregate dynastic consumption would not be constant across periods. Using this fact and applying a first-order approach to the incentive compatibility constraints (1.27), the component planning problem at a steady state can be written recursively as:

$$\begin{aligned} J(w, s) = \max_{\{c,y,w^1,w^0\}} \int_{\theta} \Big\{ & \mathbb{I}(s = 1) u(c^o(\theta)) + \hat{\beta} \left(u(c^y(\theta)) - h\left(\frac{y(\theta)}{\theta} \right) \right) \\ & - \hat{\lambda} \Big[\mathbb{I}(s = 1) c^o(\theta) + c^y(\theta) - y(\theta) \Big] + \hat{\beta}\delta \sum_{s'} \pi(s'|\theta) J(w^{s'}(\theta), s') \Big\} f(\theta) d\theta \end{aligned} \quad (1.35)$$

subject to

$$\mathcal{V}(\theta) = \mathbb{I}(s=1)u(c^o(\theta)) + \beta\left(u(c^y(\theta)) - h\left(\frac{y(\theta)}{\theta}\right)\right) + \beta\delta\sum_{s'}\pi(s'|\theta)w^{s'}(\theta), \tag{1.36}$$

$$\mathcal{V}'(\theta) = \beta h'\left(\frac{y(\theta)}{\theta}\right)\frac{y(\theta)}{\theta^2} + \beta\delta\sum_{s'}\frac{\partial\pi(s'|\theta)}{\partial\theta}w^{s'}(\theta), \tag{1.37}$$

$$w = \int_{\theta}\mathcal{V}(\theta)f(\theta)d\theta, \tag{1.38}$$

where $w^{s'}$ denotes the continuation utility contingent on the future survival state being $s' \in \{0,1\}$.

Just like in the two-period model, solutions to (1.35) satisfy the original set of incentive compatibility constraints (1.27) under certain monotonocity conditions on the optimal allocations. The next Lemma establishes this result. (The proof is omitted as it analogous to its counterpart in the two-period model in Lemma 1.2.1.) Monotonicity conditions are verified ex-post in the numerical simulations.

Lemma 1.3.1. *Suppose that for all (w,s) the solution to (1.35) satisfies:*

$$\frac{dy(\theta)}{d\theta} \geq 0, \quad \frac{dc^y(\theta)}{d\theta} \geq 0, \quad \frac{dc^o(\theta)}{d\theta} \geq 0, \quad \frac{dw(\theta)}{d\theta} \geq 0, \quad \text{and} \quad \frac{d\tilde{w}(\theta)}{d\theta} \leq 0. \tag{1.39}$$

Then the allocation $\{c,y\}$ generated by the policy functions of (1.35) is incentive compatible, i.e. it satisfies (1.27).

1.3.3 OPTIMAL WEALTH TAXES

As in the two-period version of the model, I focus on a tax implementation which features a nonlinear labor income tax schedule and linear wealth taxes which depend on effective effort histories.

Under this decentralization w_1-dynasties face the following sequence of budget constraints for all (t, s^t, θ^t):

$$\begin{aligned} c_t^y(w_1,\theta^t,s^t) + k_{t+1}(w_1,\theta^t,s^t) \leq\ & y_t(w_1,\theta^t,s^t) \\ & + (1-\mathbb{I}(s_t=1))R_t k_t(w_1,\theta^{t-1},s^{t-1})(1-t_t^b(w_1,y^t,s^t)) \\ & + \mathbb{I}(s_t=1)g_t(w_1,\theta^t,s^t)(1-t_t^i(w_1,y^t,s^t)) - T_t(w_1,y^t,s^t), \end{aligned} \tag{1.40}$$

$$\begin{aligned} c_t^o(w_1,\theta^t,(s^{t-1},1)) + g_t(w_1,\theta^t,(s^{t-1},1)) \leq\ & R_t k_t(w_1,\theta^{t-1},s^{t-1})(1-t_t^k(w_1,y^t,(s^{t-1},1))) \\ & + SS_t(w_1,y^t,(s^{t-1},1)), \end{aligned} \tag{1.41}$$

where t_t^b, t_t^i, and t_t^k are linear taxes on inheritances, inter-vivos transfers, and capital returns,

respectively, T_t is a great labor income tax schedule, SS_t are government transfers received at old age (which can be interpreted as social security benefits), g_t denotes the level of inter-vivos gifts, R_t is the pre-tax gross interest rate and k_{t+1} is the level of savings at t. Note that taxes and government transfers are functions of effective labor histories $y^t \equiv (y_1, y_2, ..., y_t)$ instead of skill shock histories θ^t (I expand on this issue below).

Definition 1.3.1. *A competitive equilibrium with taxes in the infinite horizon economy is an allocation for consumption and effective effort* $\{c_t^y(w_1, \theta^t, s^t), c_t^o(w_1, \theta^t, s^t), y_t(w_1, \theta^t, s^t)\}$, *a sequence of savings and inter-vivos gifts* $\{k_t(w_1, \theta^t, s^t), g_t(w_1, \theta^t, s^t)\}$, *a tax system* $\{t_t^b(w_1, \theta^t, s^t), t_t^i(w_1, \theta^t, s^t), t_t^k(w_1, \theta^t, s^t), T_t(w_1, \theta^t, s^t)\}$, *a sequence of social security benefits* $\{SS_t(w_1, \theta^t, s^t)\}$, *and a sequence of pre-tax interest rates* $\{R_t\}$ *such that:*

1. *The allocation for consumption, effective effort, capital, and inter-vivos gifts maximize utility* (1.25) *subject to the budget constraints* (1.40) *and* (1.41).
2. *The government's budget constraint is balanced in every period.*
3. *The sequence of resource constraints in* (1.26) *holds, so that the goods market clears.*

The next assumption is required to establish the tax implementation result. For now on, starred allocations and wedges correspond to the optimal ones.

Assumption 1.3.1. *Let* $D_t(w_1, s^t) \equiv \{y^t : \exists \theta^t \text{ such that } y_r = y_r^*(w_1, \theta^r, s^r), r = 1, ..., t\}$. *Then for all* $j = y, o$, *there exists functions* $\hat{c}_t^j : \mathcal{W}_1 \times [0, \bar{y}]^t \times S^t \to \mathbb{R}_+$ *such that*

$$\hat{c}_t^j(w_1, y^t, s^t) = c_t^{j*}(w_1, \theta^t, s^t)$$

for all (w_1, s^t) *and* $y^t \in D_t(w_1, s^t)$.

In words, the set $D_t(w_1, s^t)$ defines the effective labor histories y^t such that there is some agent that is allocated y^t at the optimum when his type is (w_1, θ^t, s^t). Assumption 1.3.1 says that optimal consumption allocations can be written in terms of effective labor histories instead of skill realizations. As it is shown next, by this property taxes can be written in terms of effective labor allocations, which are observable. The result follows Kocherlakota (2005) closely.

Proposition 1.3.1. *Let* $\{c^*, y^*\}$ *be an optimal allocation that solves* (1.30) *and satisfies Assumption 1.3.1. Then there exists a tax system* $\{t_t^b(w_1, y^t, s^t), t_t^i(w_1, y^t, s^t), t_t^k(w_1, y^t, s^t), T_t(w_1, y^t, s^t)\}$ *and a sequence of social security benefits* $\{SS_t(w_1, y^t, s^t)\}$ *such that* $\{c^*, y^*\}$ *can be implemented as a competitive equilibrium with taxes.*

Proof. See Appendix 1.7.1. □

As shown in Appendix 1.7.1, optimal wealth taxes under this implementation are given by

$$1 - t_t^i(w_1, y^t, s^t) = \frac{u'(\hat{c}_t^o(w_1, y^t, s^t))}{\beta u'(\hat{c}_t^y(w_1, y^t, s^t))}, \tag{1.42}$$

$$1 - t_t^k(w_1, y^t, (s^{t-1}, 1)) = \frac{u'(\hat{c}_{t-1}^y(w_1, y^{t-1}, s^{t-1}))}{R_t \delta u'(\hat{c}_t^o(w_1, (y^{t-1}, y_t), (s^{t-1}, 1)))}, \tag{1.43}$$

$$1 - t_t^b(w_1, y^t, (s^{t-1}, 0)) = \frac{u'(\hat{c}_{t-1}^y(w_1, y^{t-1}, s^{t-1}))}{R_t \beta \delta u'(\hat{c}_t^y(w_1, (y^{t-1}, y_t), (s^{t-1}, 0)))}, \tag{1.44}$$

for all (w_1, s^t) and $y^t \in D_t(w_1, s^t)$.

WEALTH TAXATION AT THE STEADY STATE

Given that there is no capital in this economy, the sequence of pre-tax rates $\{R_t\}$ is undetermined. Hence, the tax formulas above imply that the level of optimal capital and bequest taxes cannot be uniquely pinned down. To make progress I follow Farhi and Werning (2006) and use the interest rate that would hold at a steady state in an economy *with* capital.[24] That is, I set

$$R_t = \frac{1}{\hat{\beta}\delta} \tag{1.45}$$

for all t.

In what follows I analyze the properties of optimal wealth taxes at a steady state under (1.45). The main conclusion of the analysis is that the key properties of wealth taxes obtained in the two-period model extend to the current framework. To simplify notation, I focus on steady state wedges which are written in terms of skills rather than effective labor. This is without loss of generality as the optimal marginal wealth taxes obtained earlier map exactly to the newly defined wedges at the optimum.

I summarize intertemporal bequest and capital distortions by means of *expected* wedges. Specifically, for each (θ, w, s) and for any allocation define the *expected* bequest and capital wedges as

$$1 - \bar{\tau}^b(\theta, w, s) \equiv \frac{\hat{\beta}}{\beta} u'(c^y(\theta, w, s)) \int_{\theta'} \frac{1}{u'(c^y(\theta', w^0(\theta, w, s), 0))} dF(\theta'), \tag{1.46}$$

and

$$1 - \bar{\tau}^k(\theta, w, s) \equiv \hat{\beta} u'(c^y(\theta, w, s)) \int_{\theta'} \frac{1}{u'(c^o(\theta', w^1(\theta, w, s), 1))} dF(\theta'), \tag{1.47}$$

respectively.[25]

[24] Farhi and Werning (2006) make an analogous assumption in a dynastic environment with one-period lived agents.

[25] At any time t, the expected wedge $\bar{\tau}^b(\theta_t, w_t, s_t)$ corresponds to the expected marginal bequest tax paid by an

For each (θ, w, s), steady state inter-vivos wedges are naturally defined by

$$1 - \tau^i(\theta, w, 1) \equiv \frac{u'(c^o(\theta, w, 1))}{\beta u'(c^y(\theta, w, 1))}. \tag{1.48}$$

Now I establish the main result of the section.

Proposition 1.3.2. *For each (w, s), at any solution to the relaxed planning problem (1.32) satisfying (1.39) optimal wealth wedges satisfy*

$$\bar{\tau}^{k*}(\theta, w, s) = \beta \frac{\mu(\theta, w, s)}{f(\theta)} \frac{P'(\theta)}{P(\theta)} \frac{u'(c^{y*}(\theta, w, s))}{\hat{\lambda}}, \tag{1.49}$$

$$\bar{\tau}^{b*}(\theta, w, s) = -\hat{\beta}\left(\frac{\hat{\beta}}{\beta} - 1\right) \frac{u'(c^{y*}(\theta, w, s))}{\hat{\lambda}} - \beta \frac{\mu(\theta, w, s)}{f(\theta)} \frac{P'(\theta)}{1 - P(\theta)} \frac{u'(c^{y*}(\theta, w, s))}{\hat{\lambda}}, \tag{1.50}$$

$$\tau^{i*}(\theta, w, 1) = -\left(\frac{\hat{\beta}}{\beta} - 1\right) \frac{u'(c^{o*}(\theta, w, 1))}{\hat{\lambda}}, \tag{1.51}$$

where $\mu(\theta, w, s) \geq 0$ is the costate associated with (1.37).

Proof. See Appendix 1.7.1. □

Proposition 1.3.2 shows that the key properties of wealth wedges derived in the simpler two-period version of the model also apply to the infinite horizon environment (compare to Proposition 1.2.3). In particular, marginal bequest taxes can be decomposed into a Pigouvian and a differential mortality term, expected marginal capital taxes are positive and marginal taxes on inter-vivos transfers are negative and progressive.

There are a few notable differences with the two-period model. First, optimal taxes vary with the *current* survival state s.[26] Second, in this environment, bequest and inter-vivos taxes are not unified for two reasons. The first one is heterogeneity in mortality risk. This feature creates differences in the slopes of the wedges, as in the two-period version. But differently from the two-period model, expected capital and inter-vivos taxes do not typically coincide even in the absence of heterogeneity in life expectancies. This follows from the presence of future skill uncertainty.

individual of type (θ_t, w_t, s_t) at $t+1$. A similar interpretation holds for $\bar{\tau}^k(\theta_t, w_t, s_t)$.

[26] In Section 1.5 I show that such variation is quantitatively relevant.

1.4 CALIBRATION

1.4.1 PRELIMINARIES

I assume that individuals have constant relative risk aversion (CRRA) preferences over consumption and isoelastic disutility on effort so that

$$u(c) = \begin{cases} \frac{c^{1-\gamma}}{1-\gamma}, & \text{if } \gamma > 0,\ \gamma \neq 1, \\ \log(c), & \text{if } \gamma = 1, \end{cases} \tag{1.52}$$

where γ is the coefficient of relative risk aversion, and

$$h(l) = \frac{l^{1+\frac{1}{\varepsilon}}}{1+\frac{1}{\varepsilon}}, \tag{1.53}$$

where ε is the Frisch elasticity of labor supply.

While in the model mortality heterogeneity is only tied to the ability of the agents, for calibration purposes it will be convenient to model mortality risk as generally as possible. Hence, I assume that mortality risk is indexed by $\nu \in \mathbb{R}_+$ which represents each individual's "frailty" and summarizes all sources of mortality risk other that age.

I use a mixed proportional hazard (MPH) model for mortality risk, along the lines of Vaupel et al. (1979). Specifically, the mortality hazard rate of an individual with frailty type ν at time t is given by

$$\eta_t(\nu) = \nu \eta_t, \tag{1.54}$$

where η_t is the *baseline* mortality hazard rate at t.

As it is common practice in duration analysis, I parameterize the baseline hazard rate to make progress with the estimation of the model.[27] Following Einav et al. (2010), I assume that the date of death follows a Gompertz distribution with shape parameter ϕ. This is equivalent to imposing that the baseline hazard rate is given by

$$\eta_t = \exp(\phi \cdot t), \tag{1.55}$$

Finally, I assume that θ and ν are jointly log-normally distributed:[28]

$$\begin{pmatrix} \log(\theta) \\ \log(\nu) \end{pmatrix} \sim N\left(\begin{bmatrix} \mu_\theta \\ \mu_\nu \end{bmatrix}, \begin{bmatrix} \sigma_\theta^2 & \rho\sigma_\theta\sigma_\nu \\ \rho\sigma_\theta\sigma_\nu & \sigma_\nu^2 \end{bmatrix} \right) \tag{1.56}$$

[27] See Wooldridge (2002), Ch. 20. As an exception, in Hosseini (2014) mortality heterogeneity is identified by using data on subjective survival probabilities.

[28] Log-normally distributed skills is a standard assumption in the literature if one does not focus on the top of the income distribution (see Saez (2001) and Mankiw et al. (2009)). This specification is justified here given that data from the Health and Retirement Study does not include very rich individuals.

As discussed next, the correlation between θ and ν is a crucial input for the calibration of $P(\theta)$.

1.4.2 PROBABILITY OF SURVIVAL ACROSS SKILLS

The probability of survival across skill types is the main empirical object of interest in the analysis. A number of studies have correlated survival probabilities to *observable* socioeconomic characteristics, such as education, income or wealth. However, estimating this correlation across *unobservable* earning abilities is more rare.[29]

I calibrate the probability of survival across skills in two steps: (i) I parameterize this function based on the correlation between frailty and skills in (1.56), and (ii) I calibrate such correlation by targeting the relationship between mortality outcomes and income in U.S. data. This subsection deals with (i). Step (ii) is relegated to the following subsections.

Under (1.54) and (1.55) the probability of survival to age t from birth for a frailty type ν is given by[30]

$$\tilde{P}_t(\nu) = \exp\left(\frac{\nu}{\phi}\left(1 - \exp(\phi t)\right)\right). \tag{1.57}$$

Now define

$$P_t(\theta) \equiv \tilde{P}_t(\bar{\nu}(\theta)), \quad \text{with} \quad \bar{\nu}(\theta) \equiv \exp\left(\mathbb{E}\left[\log(\nu)|\log(\theta)\right]\right).$$

Note that $\log(\bar{\nu}(\theta))$ is the conditional mean of $\log(\nu)$ given $\log(\theta)$. Using (1.57) and the distributional assumption (1.56), $P_t(\theta)$ can be written as

$$P_t(\theta) = \tilde{P}_t\Big(\exp\left(\mu_\nu + \rho\sigma_\nu/\sigma_\theta(\log(\theta) - \mu_\theta)\right)\Big). \tag{1.58}$$

The deterministic function $P_t(\theta)$ corresponds to the probability of survival across skill types used to solve the planning problem.

1.4.3 PARAMETER CALIBRATION

Table 1.1 shows the values of the model parameters in the benchmark calibration. A set of parameters is chosen based on previous studies. Within this group, I set the annual intertemporal discount factor to 0.96, the risk aversion coefficient to 1 and the Frisch elasticity to 0.5 following Chetty et al. (2011). Each period in the model comprises 25 years. The distribution of skills is taken from Mankiw et al. (2009), who proxy ability using hourly wages in the 2007 Current Population Survey (CPS) rotating March sample.

The rest of the parameters are either estimated or calibrated to match certain features of the data. Such procedures are described in the following subsections. The main data source is the

[29] One exception is the study of Hosseini and Shourideh (2014) who work with discrete types.
[30] See Appendix 1.7.2 for details.

Health and Retirement Study (HRS), which is a panel survey is administered by the Institute for Social Research at the University of Michigan. This survey interviews individuals over 50 years old and their spouses on a biennial basis. It provides detailed information on income, assets, and mortality which makes it particularly suitable for calibrating the probability of survival across earning abilities.

Table 1.1: Calibration.

Parameter	Symbol	Value	Source
Period length	T	25 years	
Risk aversion	γ	1	
Annual subjective discount factor	$\delta^{\frac{1}{T}}$	0.96	
Frisch elasticity	ε	0.50	Chetty et al. (2011)
Mean of $\log(\theta)$	μ_θ	2.76	Mankiw et al. (2009)
Std. dev. of $\log(\theta)$	σ_θ	0.56	Mankiw et al. (2009)
Mean of $\log(\nu)$	μ_ν	-5.42	Estimation
Std. dev. of $\log(\nu)$	σ_ν	1.13	Estimation
Gompertz shape parameter	ϕ	0.09	Estimation
Correlation between $\log(\theta)$ and $\log(\nu)$	ρ	-0.14	Calibration
Altruism coefficient	β	1.34	Calibration

MORTALITY HETEROGENEITY

To estimate μ_ν, σ_ν, and ϕ I follow a standard methodology in survival analysis which uses realized mortality outcomes of a given cohort of individuals. The approach is particularly close to the one in Einav et al. (2010) who estimate analogous parameters for the U.K. based on restricted annuity data.

Mortality data in HRS for individual i can summarized by $m_i = (s_i, t_i, d_i)$, where s_i is the age when individual i entered the sample, t_i is the age when the individual exited and $d_i \in \{0, 1\}$ indicates whether the individual exited because of death ($d_i = 1$) or censoring ($d_i = 0$). The ages of entry and exit are reported in days, so I treat mortality as a continuous process. Given this information and denoting by a the actual date of death, the likelihood of observing m_i is

$$\Pr(m_i|\nu,\phi) = \Pr\left(a = t_i | t > s_i, \nu, \phi\right)^{d_i} \Pr\left(a \geq t_i | t > s_i, \nu, \phi\right)^{1-d_i}, \tag{1.59}$$

or

$$\Pr(m_i|\nu,\phi) = \left(-\frac{\partial P_{t_i}(\nu,\phi)/\partial t}{P_{s_i}(\nu,\phi)}\right)^{d_i} \left(\frac{P_{t_i}(\nu,\phi)}{P_{s_i}(\nu,\phi)}\right)^{1-d_i}, \tag{1.60}$$

where $\partial P_{t_i}(\nu,\phi)/\partial t = -\nu \exp\left(\phi t + \frac{\nu}{\phi}\left(1 - \exp(\phi t)\right)\right)$ is the Gompertz density.

The log-likelihood is

$$\mathcal{L}\left(\phi, \mu_\nu, \sigma_\nu | (m_i)_{i=1}^N\right) = \sum_{i=1}^{N} \log\left(\int \Pr(m_i|\nu, \phi) g(\nu|\mu_\nu, \sigma_\nu) d\nu\right). \tag{1.61}$$

I use mortality outcomes of households who responded to the first wave of the HRS survey in 1992, which contains individuals born between 1931-1941. I restrict my benchmark sample to male respondents between 50-64 years old in 1992. I only keep respondents who were alive and responded to the first wave and follow these individuals throughout the last wave (2010) until they either die or exit due to right-censoring. The total number of observations is 3,481 respondents, of which 1,315 died within the observed period.

Table 1.2 presents the results of the estimation under three different samples. The first row corresponds to the benchmark sample described previously. The second and third rows show the results when the sample is modified by increasing the maximum possible age to 84 years old and by including females, respectively. As expected, the mean frailty increases when older individuals are included. Besides, older respondents put a downward pressure on the degree of mortality heterogeneity. Such result is in line with the conclusions of Hurd et al. (2001), who find evidence that mortality differentials diminish with age.[31] Finally, by including females mean frailty decreases but the estimated degree of heterogeneity increases with respect to the benchmark estimates. This is consistent with the well documented fact that women on average live longer than men do.

Table 1.2: Mortality Parameter Estimates.

Sample	Estimates			No. of Obs.	% of Deaths
	μ_ν	σ_ν	ϕ		
50-64 males	-5.416	1.133	0.088	3,481	37.78
	(0.391)	(0.278)	(0.016)		
50-84 males	-5.266	0.930	0.085	3,790	40.18
	(0.276)	(0.226)	(0.011)		
50-64 males and females	-6.099	1.416	0.104	7,211	31.34
	(0.217)	(0.100)	(0.008)		

Notes: The three samples are drawn from the initial HRS survey in 1992. The first row corresponds to the benchmark sample. Ages are measured at the beginning of the first wave. "% of Deaths" denotes the fraction of individuals who died within each sample between waves 1 and 10. Standard errors are reported in parentheses.

[31] Their estimation is based on mortality outcomes between waves 1 and 2 of the Asset and Health Dynamics among the Oldest-Old (AHEAD) study, a sub-sample of households in HRS born in 1923 or earlier.

CORRELATION BETWEEN MORTALITY AND SKILLS

I calibrate ρ, the correlation between $\log(\theta)$ and $\log(\nu)$, by matching the relationship between mortality and permanent income in HRS data. Details follow.

I center attention on mortality rates for male retirees between 65-75 years old across waves 1998 and 2000. The focus on retirees is justified on two grounds. First, the majority of individuals who die in a given wave were previously retired. Second, HRS provides detailed information on retirement income which can be used to stratify the population into permanent income categories (see below). I use mortality outcomes between waves 1998-2000 only because it maximizes the number of deaths across waves given my sample restrictions.

For each respondent, *permanent* income is measured as the average of *current* non-asset retirement income over waves 1998 and 2000, as long as the individual is alive. Current non-asset income is the sum of income from Social Security retirement benefits, employer pensions, annuities, veteran's benefits, welfare, and food stamps. My measure of permanent income is similar to the one in De Nardi et al. (2010), which is motivated by the fact that Social Security and pension benefits are typically increasing functions of labor income before retirement. I define permanent income quartiles separately for singles and couples using sample weights provided by HRS. The sample consists of 2,989 respondents, of which 450 died between waves.[32]

The first column of Table 1.3 shows the two-year weighted mortality rates profiles across permanent income quartiles in the data. I calibrate ρ to match those moments to its simulated counterparts.[33] In a nutshell, the artificial data is generated by simulating the life spans of a large number of individuals indexed by wage-age pairs. The second column of the table reports the moments generated by the calibrated value of ρ which is -0.14. Appendix 1.7.2 provides additional details.

Table 1.3: Two-year Mortality Rates: Model Vs. Data

Income Quartile	Data	Model
Lowest	.137	.132
2	.077	.098
3	.059	.088
Highest	.056	.049

Notes: "Data" shows weighted mortality rates between waves 1998 and 2000 in HRS for retirees between 65-75 years old. "Model" corresponds to the moments generated by the numerical simulations.

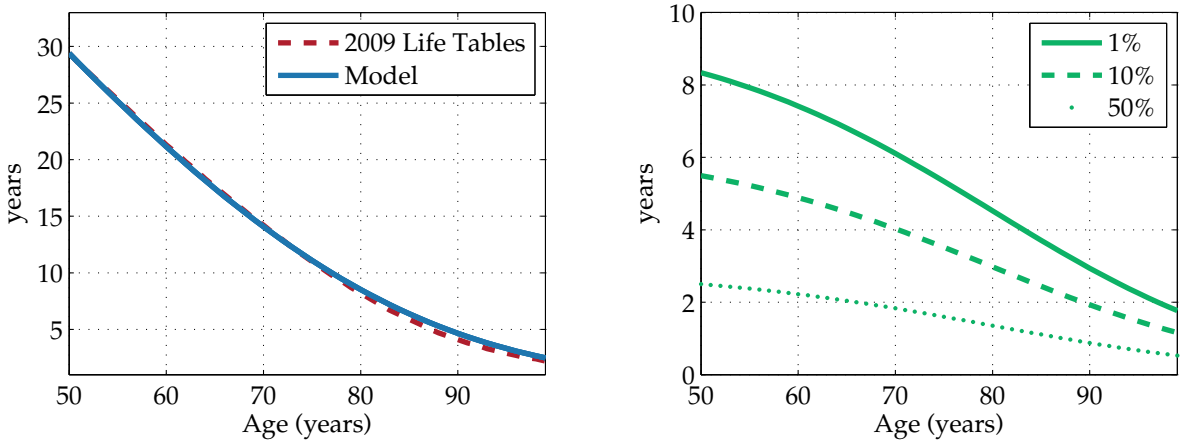
Figure 1.1 illustrates some properties of the average life expectancies implied by the calibrated

[32] Individuals with zero or missing values for permanent income are dropped.

[33] Hosseini (2014) applies a similar method to estimate the relationship between earnings and mortality.

probability of survival $P_t(\theta)$ using (1.58).[34] As a measure of out-of-sample fit, panel (a) compares life expectancies across ages in the model to those in the 2009 period life table for U.S. males.[35] I plot life expectancies for individuals over 50 years old, which corresponds to the age group in the HRS survey. It is worth noting that the model fits the data very closely along this dimension, in particular for individuals younger than 80.

Panel (b) plots the differences in average life expectancies between certain top and bottom percentiles in the model. Notice that individuals of 50 years old at the top 1% of the ability distribution are expected to live almost 8.5 years more than individuals at the bottom 1%, which provides evidence of substantial mortality differences across skills. It is useful to compare these numbers to previous studies. Waldron (2007), for example, estimates that the difference in life expectancies at 65 for male social security-covered workers between the top and bottom halves of the income distribution is around 1.9 years for the period 1999-2000.[36] The counterpart in this model is around 2.04 years.



(a) Comparing with Life Tables.

(b) Top minus Bottom.

Figure 1.1: Average Life Expectancies across Ages. Panel (a) compares average life expectancies in the model to those in the 2009 period life tables for U.S. males. Panel (b) plots the differences in average life expectancies between top and bottom 1%, 10% and 50% of the skill distribution predicted by the model.

[34] The average expectation of life at age t is given by $e_t \equiv \int_\Theta \int_t^\infty \frac{P_s(\theta)}{P_t(\theta)} ds d\theta$.

[35] See http://www.cdc.gov.

[36] She uses the Social Security Administration's Continuous Work History Sample. This is a much larger sample than the one used in this paper, but it is not publicly available.

COEFFICIENT OF ALTRUISM

While many studies have tested whether altruistic bequest motives are present in the data, to the best of my knowledge few provide an estimate of the altruism coefficient when the null hypothesis holds.[37] One exception can be found in the work of Abel and Warshawsky (1988), who calibrate the altruism coefficient to be consistent with a steady state in a deterministic economy without taxes. I calibrate β by applying their methodology in this framework. Specifically, it is straightforward to show that a steady state in a deterministic economy without taxes is only consistent with $\beta = (\delta R)^{-1}$. Using the calibrated value of δ in Table 1.1 and an annual interest rate of 3%, this formula yields $\beta = 1.33$.[38]

This methodology is admittedly rough. However, a similar value for β can be obtained by applying an alternative calibration strategy that relies on consumption data.[39] Specifically, consider the decentralized version of the infinite horizon economy presented in Section 1.3.3. Assuming that taxes on inter-vivos transfers are zero, one gets the intergenerational Euler equation:

$$u'(c_t^o(w_1, \theta^t, s^t)) = \beta u'(c_t^y(w_1, \theta^t, s^t)),$$

for all t and for all (w_1, θ^t, s^t).

Using the CRRA preferences in (1.52), this expression can be rearranged as

$$\beta = \left(\frac{C_t^y}{C_t^o}\right)^{\gamma}, \tag{1.62}$$

where $C_t^j \equiv \mathbb{E}\left[c_t^j(w_1, \theta^t, s^t)\right]$ for $j = y, o$, and the expectation is taken over (w_1, θ^t, s^t).

For a given value of γ, equation (1.62) can be used to recover the coefficient of altruism β using cross-sectional consumption data of different cohorts. I use data from the Consumption Expenditure Survey (CEX), which provides comprehensive measures of consumption for a representative cross section of households in the US on a quarterly basis. I focus on the measure of total consumption elaborated by Krueger and Perri (2006).[40] Young households are identified as those with working heads between 25 and 54 years old receiving positive labor income. The old include all households with retired heads over 65 years old. Each household in CEX is interviewed for a maximum of four times and I measure yearly consumption as the sum of the quarterly measures of consumption reported in each of these interviews. Figures are expressed in adult equivalent units in 1982-1984 constant dollars using CPI.[41] The total number of observations is 37,789, of

[37] Two examples in the first group of studies are Altonji et al. (1997) or Abel and Kotlikoff (1994).

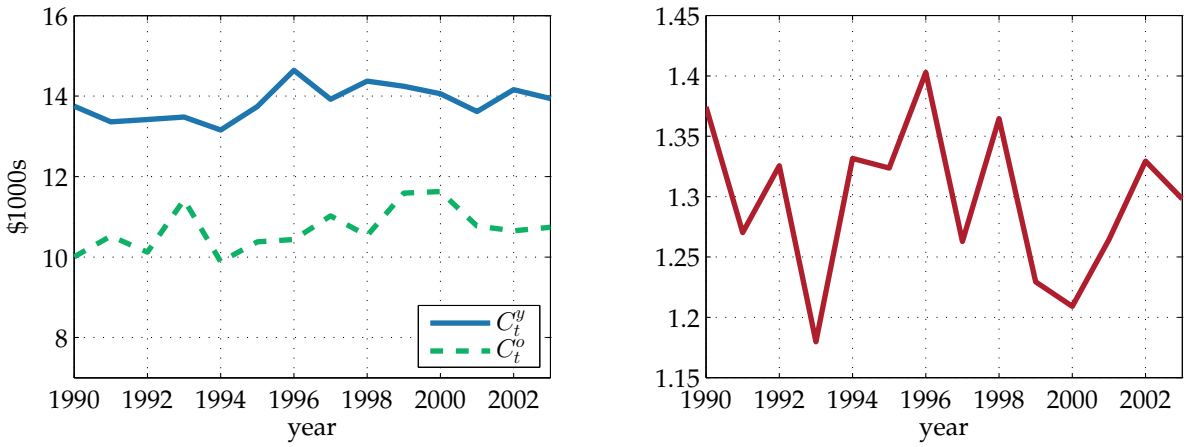
[38] Recall that the frequency in the model is 25 years.

[39] Just as in Abel and Warshawsky (1988), this method assumes that intergenerational altruism is present in the data. This is a strong assumption, as the evidence for altruism is inconclusive: Tomes (1981) and Becker and Tomes (1986), for instance, present evidence in support of the altruism hypothesis, while Altonji et al. (1992) reject some of the empirical implications of the altruistic framework.

[40] This measure of consumption includes nondurables, services and imputed values for large durables (such as housing and vehicles). For details refer to Krueger and Perri (2006).

[41] All data is weighted using CEX population weights. I exclude rural households, households who haven't

which 32,834 are young households.



(a) Average Consumption.

(b) Young-to-Old Consumption Ratio.

Figure 1.2: Consumption Data. The consumption measure corresponds to the one elaborated by Krueger and Perri (2006) using CEX data. Figures expressed in adult equivalent units in 1982-1984 constant dollars using CPI.

Figure 1.2 shows the average consumption measure across different cohorts between 1990-2003. Clearly, consumption of the young is significantly larger than for the old. Also, the ratio between these measures seems to follow a stationary process throughout the observed period. Table 1.4 reports the corresponding estimates of β consistent with (1.62). These are obtained by taking time series averages for different values of risk aversion.

As anticipated, the values of β calibrated in this fashion are fairly close to the 1.33 figure in the benchmark calibration, especially for low values of risk aversion.

Table 1.4: Calibration of β using CEX Consumption Data.

	$\gamma = 1$	$\gamma = 1.5$	$\gamma = 2$
β estimate	1.30	1.48	1.69

Notes: "β estimates" corresponds to time series averages of the coefficient of altruism implied by equation (1.62) using the annual consumption measure described in the text between 1990-2003.

completed four consecutive interviews, and observations corresponding to young households with negative after-tax labor income or those who report positive labor income but zero hours worked.

1.5 QUANTITATIVE RESULTS

Constrained optimal allocations are computed numerically as follows. First, I solve the component planning problem (1.35) by value function iteration over a grid of (w, s) for a given shadow price $\hat{\lambda}$. The value function is interpolated using Chebyshev polynomials. The skill distribution is truncated at an hourly wage of 60 dollars, which roughly corresponds to the 99th percentile of the wage distribution.[42] Given the solution of the component planning problem, I approximate a steady state distribution of continuation utility entitlements via Monte Carlo simulation. Finally, I iterate on the shadow price $\hat{\lambda}$ until the resource constraint (1.26) holds at the steady state. Optimal wedges at the steady state are interpolated using splines. Additional details on this numerical procedure are contained in Appendix 1.7.3. I solve the model for two values of the coefficient of social altruism $\hat{\beta}$; namely, 1.35 and 1.6. According to (1.45), these choices would yield annual interest rates of around 2.90% and 2.25% at the steady state in an economy with capital, respectively.

Recall that mortality differences primarily shape the optimal tax code via capital taxes and through the differential mortality term on bequest taxes. Table 1.5 reports the corresponding expected distortions (see Section 1.3.3) on the *median worker*, who is defined as the dynast who draws the median skill and receives the median continuation utility at the steady state.

Table 1.5: Expected Distortions on the Median Worker.

	$s = 1$		$s = 0$	
	$\hat{\beta} = 1.35$	$\hat{\beta} = 1.6$	$\hat{\beta} = 1.35$	$\hat{\beta} = 1.6$
Capital Wedge	1.71%	1.71%	1.27%	1.28%
DM term on Bequest Wedge	-3.39%	-3.39%	-2.52%	-2.53%

Notes: For $w = \{k, b\}$, the expected wedge on the median worker is given by $\bar{\tau}^w(Me(\theta), Me(w), s)$, where $Me(x)$ denotes the median of x.

Note that the expected capital wedge ranges between 1.3% and 1.7% and the expected bequest wedge between -3.4% and -2.5%. Notably, these magnitudes do not change much with $\hat{\beta}$. In fact, the change in capital wedges and the mortality term on bequest wedges is almost negligible when increasing the social level of altruism by about 20%. This suggest that the quantitative effect of mortality differences on optimal tax rates is quite robust to the normative criterion at hand.

The magnitudes of the distortions on capital and bequests emerging from differences in longevity are modest not negligible. For example, the expected capital tax is around 0.1% at an annual frequency. This value is commensurate with the magnitude of taxes on net worth in certain developed countries; Switzerland, for instance, imposes progressive wealth taxes which range between 0.03% and 0.94% at an annual frequency depending on the cantons.[43]

[42] Recall that the focus of the analysis is not on the tax properties at the top of the skill distribution. Under this calibration, the median worker earns around 18 dollars per hour.

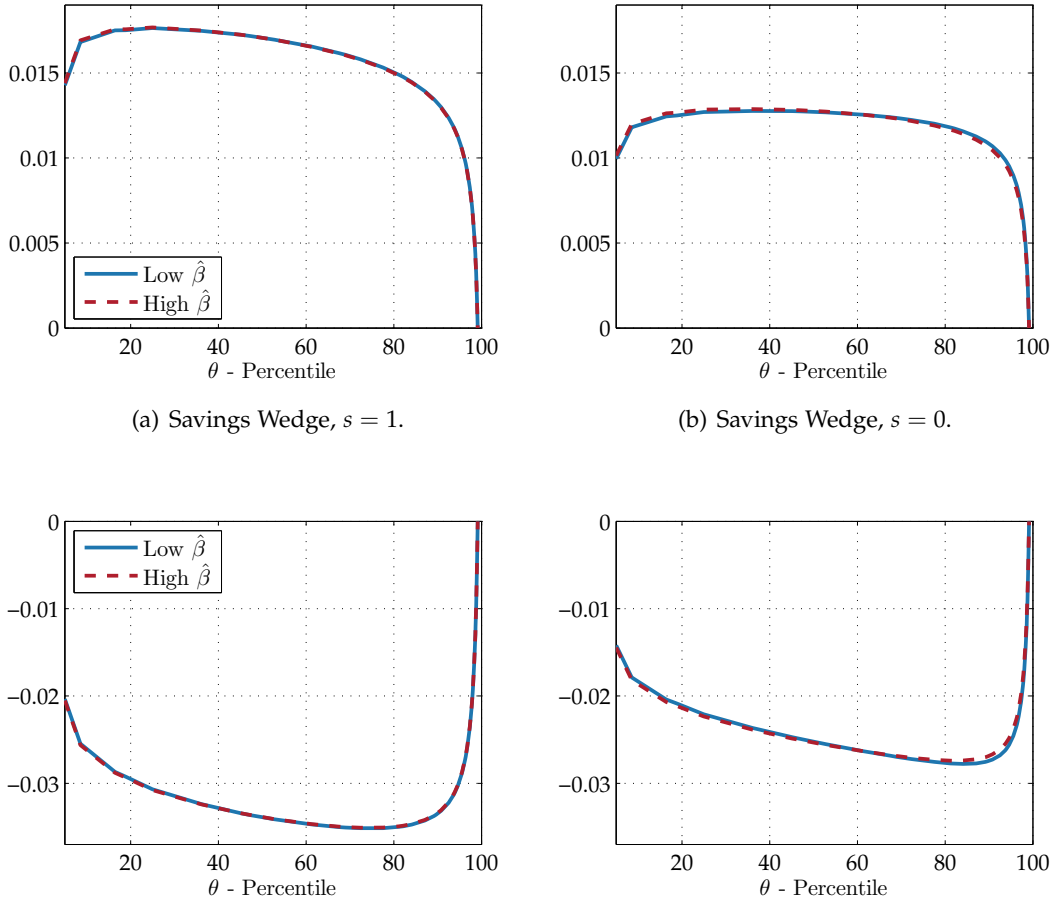
[43] See http://www.expatica.com/ch/finance/tax.

Figures 1.3 and 1.4 display expected distortions over the entire range of the skill distribution.[44] Two important lessons previously mentioned regarding the median worker also apply to the entire skill set. First, mortality-driven distortions (the capital wedge and differential mortality term on bequest wedges) are rather insensitive to the value of $\hat{\beta}$, while overall bequest and inter-vivos distortions embedding Pigouvian forces vary substantially. Second, optimal taxes can change considerably with the current survival state of the dynasty.

An interesting lesson from the simulations is that optimal distortions can vary substantially with the current survival state of the dynasty s. Expected bequest taxes, for example, can change by as much as 20 percentage points depending on whether old agents in the dynasty are alive or not. Also note that the expected distortions on capital, bequests and labor are all larger in absolute value when $s = 1$. The phenomenon can be explained by the fact that these are all distortions on *young* agents, whose relative Pareto weights falls when *old* agents are alive. This finding suggest that partial reforms based on tying the tax code to survival histories can potentially yield significant welfare gains.

Regarding the shapes of taxes on intergenerational transfers, Figure 1.4 reveals that the degree of progressivity of inter-vivos and bequest taxes changes at the high and low ends of the skill distribution: marginal inter-vivos (bequest) taxes are more progressive at the bottom (top). This is just a reflection of the U-shaped pattern of the differential mortality term affecting optimal bequest wedges.

[44] The top individual always faces zero marginal distortions due to right truncation. A common practice in public finance is to append a Pareto tail at the top of the distribution, in which case non-zero marginal wedges are obtained asymptotically (see, e.g., Diamond and Saez (2011)).



(a) Savings Wedge, $s = 1$.

(b) Savings Wedge, $s = 0$.

(c) DM Term on Bequest Wedge, $s = 1$.

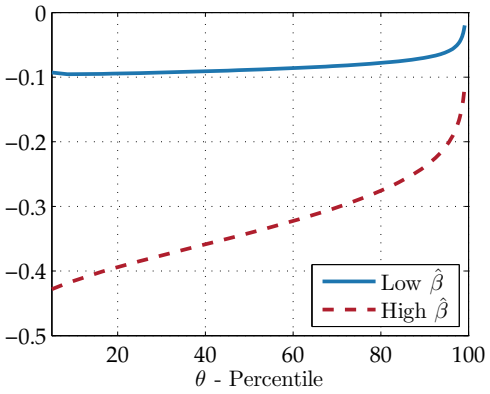
(d) DM Term on Bequest Wedge, $s = 0$.

Figure 1.3: Expected Distortions for Median w at the Steady State.

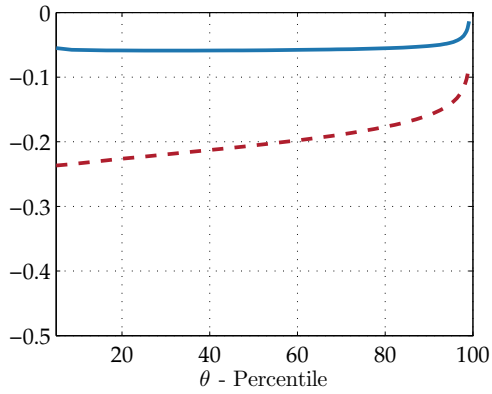
1.5.1 WELFARE ANALYSIS

In this subsection I evaluate the potential gains from tagging mortality differences across earning abilities using wealth taxes. An ideal upper bound for this exercise would be the case in which optimal wealth taxes *only* respond to the mortality gradient. As the previous analysis shows, this would correspond to a scenario where societal and private levels of altruism coincide, so that all Pigouvian drivers behind optimal taxes are shut down. This configuration encounters a serious problem, though, in the infinite horizon version of the model: in that case, no steady state exists and the standard "immiseration" result holds (refer to section 1.3.2).

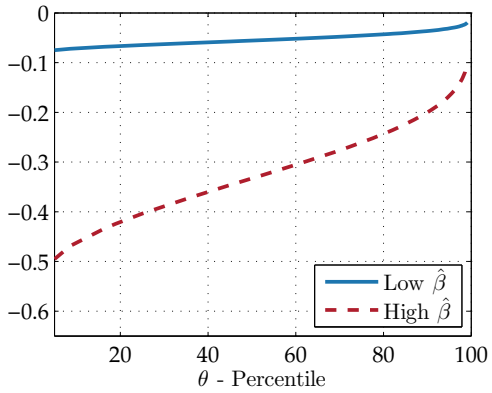
To circumvent this problem, I compute welfare gains in the two-period version of the model. These are shown in Table 1.6 for different values of Frisch elasticities and risk aversion. For the computation, I set $\hat{\beta} = \beta$ and compare welfare achieved at the full optimum and at a planning problem with two additional constraints requiring that bequest and capital wedges be *zero*. I



(a) Bequest Wedge, $s = 1$.



(b) Bequest Wedge, $s = 0$.



(c) Inter-vivos Wedge.

Figure 1.4: Expected Distortions for Median w at the Steady State.

assume an annual interest rate of 3%. The gains correspond to the factor by which consumption of all agents in all histories has to increase in the framework without bequest and capital taxes to yield the same level of welfare than in the full optimum.

Gains range between 0.015% and 0.020% of aggregate consumption, or between 1.5 and 2 billion dollars.[45] The values are small, but not negligible. To put these numbers in perspective, Golosov et al. (2013) evaluate welfare gains from using optimal capital taxes on the grounds of heterogeneous savings tastes, also in a two-period model. Their findings, however, point to negligible gains in the order of 0.00002% of aggregate consumption.

Table 1.6: Welfare Gains (in % of consumption)

	$\gamma = 1$	$\gamma = 1.5$	$\gamma = 3$
$\varepsilon = 0.50$	0.015	0.017	0.018
$\varepsilon = 0.75$	0.018	0.020	0.020

Notes: Welfare gains are computed in the two-period version of the model. ε is the Frisch elasticity of labor supply and γ is the coefficient of relative risk aversion.

1.6 CONCLUSION

Economists in the 80s initiated a profound debate around the motives behind wealth accumulation and on the relative magnitudes of the three sources of wealth, i.e. earned, inherited, and coming from inter-vivos transfers.[46] On this "savings puzzle" Laurence Kotlikoff said:

> "The answer to the savings puzzle has many policy implications; certain tax structures are much more conducive to some types of savings than others..." (Kotlikoff (1988), page 41)

This paper studies the optimal design of such "tax structures." It distinguishes between the optimal tax treatment of the three possible sources of wealth, and points to the socioeconomic mortality gradient as a crucial determinant both from a theoretical and from a quantitative angle.

The analysis admits a number of extensions. First, mortality risk was assumed to be exogenous, while in reality human and health capital shape life expectancy and interact with the tax code. Another natural step is to adapt the current framework to optimal social security design. Finally, this study recommends tying the tax system to parental survival as an optimal policy, so quantitatively evaluating partial reforms based on this feature looks like a promising route.

[45] Aggregate consumption in the U.S. economy is in the order of 10,000 billion dollars.

[46] See the exchange between Kotlikoff and Summers (1981) and Modigliani (1988). More recently, the debate was revived by Piketty (2014).

1.7 APPENDIX

1.7.1 PROOFS

PROOF OF LEMMA 1.2.1

Let $M(\theta';\theta) \equiv u(c_1^y(\theta')) - h\left(\frac{y_1(\theta')}{\theta}\right) + \delta P(\theta) w_2(\theta') + \delta(1-P(\theta))\tilde{w}_2(\theta')$. Incentive compatibility requires that for all $\theta \in \Theta$ $M(\theta';\theta)$ attain a global maximum at $\theta' = \theta$. First note that at a local maximum one must have $M_1(\theta;\theta) = 0$ and $M_{11}(\theta;\theta) \leq 0$. Using the definition of M, the first order condition can be written as

$$u'(c_1^y(\theta))\frac{dc_1^y(\theta)}{d\theta'} - h'\left(\frac{y_1(\theta)}{\theta}\right)\frac{1}{\theta}\frac{dy_1(\theta)}{d\theta'} + \delta P(\theta)\frac{dw_2(\theta)}{d\theta'} + \delta(1-P(\theta))\frac{d\tilde{w}_2(\theta)}{d\theta'} = 0. \tag{1.63}$$

Differentiating the first order condition $M_1(\theta;\theta) = 0$ with respect to θ gives $M_{11}(\theta;\theta) = -M_{12}(\theta;\theta)$. Hence, the second order condition for local maxima at the $\theta' = \theta$ is equivalent to $M_{12}(\theta;\theta) \geq 0$, or

$$\frac{dy_1(\theta)}{d\theta}\frac{1}{\theta^2}\left[h'\left(\frac{y_1(\theta)}{\theta}\right) + h''\left(\frac{y_1(\theta)}{\theta}\right)\frac{y_1(\theta)}{\theta}\right] + \delta P'(\theta)\left(\frac{dw_2(\theta)}{d\theta'} - \frac{d\tilde{w}_2(\theta)}{d\theta'}\right) \geq 0. \tag{1.64}$$

Clearly, (1.11) implies (1.64). Therefore, (1.11) and the local incentive constraints (1.9) and (1.10) guarantee that $\theta' = \theta$ is a local maximum of $M(\theta';\theta)$. I now show that the same holds for global maxima.

Evaluating (1.63) at θ' gives

$$u'(c_1^y(\theta'))\frac{dc_1^y(\theta')}{d\theta'} = h'\left(\frac{y_1(\theta')}{\theta'}\right)\frac{1}{\theta'}\frac{dy_1(\theta')}{d\theta'} - \delta P(\theta')\frac{dw_2(\theta')}{d\theta'} - \delta(1-P(\theta'))\frac{d\tilde{w}_2(\theta')}{d\theta'}.$$

Using this expression and the definition of $M(\theta';\theta)$ it follows that

$$M_1(\theta';\theta) = \frac{dy_1(\theta')}{d\theta'}\left[h'\left(\frac{y_1(\theta')}{\theta'}\right)\frac{1}{\theta'} - h'\left(\frac{y_1(\theta')}{\theta}\right)\frac{1}{\theta}\right] + \delta\left[(P(\theta)-P(\theta'))\left(\frac{dw_2(\theta')}{d\theta'} - \frac{d\tilde{w}_2(\theta')}{d\theta'}\right)\right]. \tag{1.65}$$

Now take any $\theta' < \theta$. By (1.11) the terms in square brackets in (1.65) so $M_1(\theta';\theta) \geq 0$. Analogously, for any θ' such that $\theta' > \theta$ one has that $M_1(\theta';\theta) \leq 0$. Hence, since $M_1(\theta;\theta) = 0$ I obtain

$$sign\left(M_1(\theta';\theta)\right) = sign\left(\theta - \theta'\right),$$

which implies that a global maximum is attained at $\theta' = \theta$.

PROOF OF PROPOSITION 1.2.1

Let $Y_1^* \equiv \{y_1 \in [0, \bar{y}_1] : \exists \theta \text{ such that } y_1 = y_1^*(\theta)\}$. The set Y_1^* contains the optimal allocations of effective labor. I start by establishing a preliminary result.

Lemma 1.7.1. *Suppose* (1.11) *holds. Then there exists increasing functions*

$$\hat{c}_1^y : Y_1^* \to \mathbb{R}_+, \quad \hat{\tilde{c}}_2^y : Y_1^* \to \mathbb{R}_+, \quad \hat{c}_2^j : Y_1^* \to \mathbb{R}_+,$$

such that for all θ

$$\hat{c}_1^y(y_1^*(\theta)) = c_1^{y*}(\theta), \quad \hat{\tilde{c}}_2^y(y_1^*(\theta)) = \tilde{c}_2^{y*}(\theta), \quad \hat{c}_2^j(y_1^*(\theta)) = c_2^{j*}(\theta),$$

for $j = o, y$.

Proof. The existence of the function $\hat{c}_1^y$ follows directly from (1.11). The remaining functions exist because at the optimum consumption allocations are increasing functions of each other for all types (this follows by rearranging the first order conditions of (1.8); see (1.74)-(1.77) in Appendix 1.7.1). □

Lemma 1.7.1 allows me to write optimal taxes in terms of observable effective effort rather than skills. Now define wealth taxes by

$$1 - t_2^{k*}(y_1) = \begin{cases} \frac{u'(\hat{c}_1^y(y_1))}{\delta R u'(\hat{c}_2^o(y_1))}, & \text{if } y_1 \in Y_1^*, \\ 0, & \text{otherwise,} \end{cases} \tag{1.66}$$

$$1 - t_2^{b*}(y_1) = \begin{cases} \frac{u'(\hat{c}_1^y(y_1))}{\delta R \beta u'(\hat{\tilde{c}}_2^y(y_1))}, & \text{if } y_1 \in Y_1^*, \\ 0, & \text{otherwise,} \end{cases} \tag{1.67}$$

$$1 - t_2^{i*}(y_1) = \begin{cases} \frac{u'(\hat{c}_2^o(y_1))}{\beta u'(\hat{\tilde{c}}_2^y(y_1))}, & \text{if } y_1 \in Y_1^*, \\ 0, & \text{otherwise.} \end{cases} \tag{1.68}$$

Note that for $y_1 \in Y_1^*$, $t_2^{w*}(y_1(\theta)) = \tau^{w*}(\theta)$ with $w = \{k, b, i\}$, where $\{\tau^{w*}(\theta)\}_{w=k,b,i}$ are the wedges defined in (1.12)-(1.14) evaluated at the optimal allocation.

Next I define income taxes and transfers in the optimal system. Before doing so, it is worth noting that these quantities can only be pinned down for a given prescribed plan for savings and inter-vivos gifts. Put differently, the levels of savings, gifts, income taxes and transfers are undetermined in the decentralization. This follows by applying a standard Ricardian equivalence argument. In my notion of implementation, I consider the case in which savings and inter-vivos transfers are constant across types, but many other choices would work. So suppose that all agents are induced to save the per capita amount K_2 and make inter-vivos gifts in the amount G_2. Then define the income tax T_1 and transfers SS_2 and $\tilde{SS}_2$ as

$$T_1^*(y_1) = \begin{cases} y_1 + W_1 - \hat{c}_1^y(y_1) - K_2, & \text{if } y_1 \in Y_1^*, \\ +\infty, & \text{otherwise,} \end{cases} \tag{1.69}$$

$$SS_2^*(y_1) = \begin{cases} \hat{c}_2^o(y_1) + G_2 - RK_2(1 - t_2^{k*}(y_1)), & \text{if } y_1 \in Y_1^*, \\ -\infty, & \text{otherwise,} \end{cases} \tag{1.70}$$

$$\tilde{SS}_2^*(y_1) = \begin{cases} \hat{\tilde{c}}_2^y(y_1) - RK_2(1 - t_2^{b*}(y_1)), & \text{if } y_1 \in Y_1^*, \\ -\infty, & \text{otherwise,} \end{cases} \tag{1.71}$$

Consider the subproblem of a θ-dynasty consisting on choosing consumption allocations for a given level of effective effort y_1. The solution to this problem is characterized by the Euler equations

$$u'(c_1^y(\theta)) = \delta R\Big[P(\theta)u'(c_2^o(\theta))(1 - t_2^{k*}(y_1)) + (1 - P(\theta))\beta u'(\tilde{c}_2^y(\theta))(1 - t_2^{b*}(y_1))\Big], \tag{1.72}$$

and

$$u'(c_2^o(\theta)) = \beta u'(c_2^y(\theta))(1 - t_2^{i*}(y_1)), \tag{1.73}$$

together with the budget constraints (1.17)-(1.20).

These conditions are satisfied at $\{\hat{c}_1^y(y_1), \hat{c}_2^o(y_1), \hat{c}_2^y(y_1), \hat{\tilde{c}}_2^y(y_1), K_2, G_2\}$ for all $y_1 \in Y_1^*$, given the taxes and transfers defined in (1.66)-(1.71). Note that agents would never choose $y_1 \notin Y_1^*$ due to large penalties.

Finally, consider the subproblem of choosing y_1 given the optimal choices of consumption previously characterized. To complete the proof, I need to show that each θ-type chooses $y_1^*(\theta)$. But this simply follows because the optimal allocation is incentive compatible and by applying the definition of the consumption functions in Lemma 1.7.1. The goods market clears given that the optimal allocation is resource feasible.

PROOF OF PROPOSITION 1.2.3

After integrating by parts, the Lagrangian to the planner's problem (1.8) can be written as

$$\begin{aligned}\mathcal{L} = &\int \Big[u(c_1^y(\theta)) - h\left(\frac{y_1(\theta)}{\theta}\right) + P(\theta)\delta\Big(u(c_2^o(\theta)) + \hat{\beta}u(c_2^y(\theta))\Big) + (1-P(\theta))\delta\hat{\beta}u(\tilde{c}_2^y(\theta))\Big]dF(\theta)\\ &+\lambda\int\Big[y_1(\theta) - c_1^y(\theta) - \frac{P(\theta)}{R}\Big(c_2^o(\theta) + c_2^y(\theta)\Big) - \frac{1-P(\theta)}{R}\tilde{c}_2^y(\theta)\Big]dF(\theta)\\ &+\int\gamma\Big[u(c_1^y(\theta)) - h\left(\frac{y_1(\theta)}{\theta}\right) + P(\theta)\delta\Big(u(c_2^o(\theta)) + \beta u(c_2^y(\theta))\Big) + (1-P(\theta))\delta\beta u(\tilde{c}_2^y(\theta)) - \mathcal{V}(\theta)\Big]d\theta\\ &+\int\Big[\mu'\mathcal{V}(\theta) + \mu\Big(h'\left(\frac{y_1(\theta)}{\theta}\right)\frac{y_1(\theta)}{\theta^2} + P'(\theta)\delta\left(u(c_2^o(\theta)) + \beta u(c_2^y(\theta)) - \beta u(\tilde{c}_2^y(\theta))\right)\Big)\Big]d\theta.\end{aligned}$$

For each θ, first order conditions for consumption allocations are

$$u'(c_1^y(\theta))(f(\theta)+\gamma(\theta)) = \lambda f(\theta), \tag{1.74}$$

$$u'(c_2^o(\theta))\left(P(\theta)\delta f(\theta) + \gamma P(\theta)\delta - \mu(\theta)P'(\theta)\delta\right) = \lambda\frac{P(\theta)}{R}f(\theta). \tag{1.75}$$

$$u'(c_2^y(\theta))\left(P(\theta)\delta\hat{\beta}f(\theta) + \gamma P(\theta)\delta\beta - \mu(\theta)P'(\theta)\delta\beta\right) = \lambda\frac{P(\theta)}{R}f(\theta), \tag{1.76}$$

$$u'(\tilde{c}_2^y(\theta))\left(\delta(1-P(\theta))\hat{\beta}f(\theta) + \gamma(1-P(\theta))\delta\beta + \mu(\theta)P'(\theta)\delta\beta\right) = \lambda\frac{1-P(\theta)}{R}f(\theta). \tag{1.77}$$

Rearranging (1.74) and (1.75) and applying the definition of τ^k gives (1.22). The expressions for bequest and inter-vivos wedges follow, respectively, from combining (1.74)-(1.77) and (1.75)-(1.76) and using the corresponding definitions of the wedges.

To show that $\lambda > 0$ and $\mu(\theta) \geq 0$, first note that the first order condition with respect to the state $\mathcal{V}(\theta)$ gives

$$\mu'(\theta) = -\gamma(\theta). \tag{1.78}$$

Combining (1.74) and (1.78) gives

$$\mu'(\theta) = f(\theta) - \lambda\frac{1}{u'(c_1^y(\theta))}f(\theta). \tag{1.79}$$

Integrating over $\left[\underline{\theta}, \bar{\theta}\right]$ gives

$$\frac{1}{\lambda} = \int\frac{1}{u'(c_1^y(\theta'))}f(\theta')d\theta'. \tag{1.80}$$

Equation (1.80) and Inada conditions imply that $\lambda > 0$.

Integrating (1.79) between $[\theta, \bar{\theta}]$ yields

$$\frac{\mu(\theta)}{\lambda} = \int_{\theta}^{\bar{\theta}} \frac{1}{u'(c_1^y(\theta'))} f(\theta') d\theta' - (1 - F(\theta)) \int \frac{1}{u'(c_1^y(\theta'))} f(\theta') d\theta'. \tag{1.81}$$

Under (1.11), the right hand side of (1.81) is positive, so that $\mu(\theta) \geq 0$ as well.

PROOF OF PROPOSITION 1.3.1

The implementation proof is very similar to the one in Kocherlakota (2005) and it uses many of the insights of the two-period version of the model. Hence, here I only provide a sketch.

For each (w_1, s^t), define wealth taxes by

$$1 - t_t^i(w_1, y^t, s^t) = \begin{cases} \frac{u'(\hat{c}_t^o(w_1, y^t, s^t))}{\beta u'(\hat{c}_t^y(w_1, y^t, s^t))}, & \text{if } y^t \in D_t(w_1, s^t), \\ 0, & \text{otherwise,} \end{cases} \tag{1.82}$$

$$1 - t_t^k(w_1, y^t, (s^{t-1}, 1)) = \begin{cases} \frac{u'(\hat{c}_{t-1}^y(w_1, y^{t-1}, s^{t-1}))}{R_t \delta u'(\hat{c}_t^o(w_1, (y^{t-1}, y_t), (s^{t-1}, 1)))}, & \text{if } y^t \in D_t(w_1, s^t), \\ 0, & \text{otherwise,} \end{cases} \tag{1.83}$$

$$1 - t_t^b(w_1, y^t, (s^{t-1}, 0)) = \begin{cases} \frac{u'(\hat{c}_{t-1}^y(w_1, y^{t-1}, s^{t-1}))}{R_t \beta \delta u'(\hat{c}_t^y(w_1, (y^{t-1}, y_t), (s^{t-1}, 0)))}, & \text{if } y^t \in D_t(w_1, s^t), \\ 0, & \text{otherwise,} \end{cases} \tag{1.84}$$

In this implementation agents are induced to accumulate zero capital and leave an arbitrary amount of gifts. Labor income taxes and social security transfers are chosen so that budget constraints hold with equality when $y^t \in D_t(w_1, s^t)$ and impose large penalties otherwise.

The tax choices work because the agent's consumption Euler equations are satisfied for all (w_1, s^t) and any $y^t \in D_t(w_1, s^t)$. Incentive compatibility ensures that agents choose the effective labor allocation intended for them by the planner. Markets clear under the optimal allocation by construction.

PROOF OF PROPOSITION 1.3.2

The Lagrangian to the relaxed planning problem in (1.35) is

$$\begin{aligned}\mathcal{L} = &\int \Big\{\mathbb{I}(s=1)u(c^o(\theta)) + \hat{\beta}\left(u(c^y(\theta)) - h\left(\frac{y(\theta)}{\theta}\right)\right) - \hat{\lambda}\left[\mathbb{I}(s=1)c^o(\theta) + c^y(\theta) - y(\theta)\right] \\ &+ \hat{\beta}\delta P(\theta)J(w^1(\theta),1) + \hat{\beta}\delta(1-P(\theta))J(w^0(\theta),0)\Big\}dF(\theta) \\ &+ \phi\Big[w - \int \Big\{\mathbb{I}(s=1)u(c^o(\theta)) + \beta\left(u(c^y(\theta)) - h\left(\frac{y(\theta)}{\theta}\right)\right) + \beta\delta P(\theta)w^1(\theta) + \beta\delta(1-P(\theta))w^0(\theta)\Big\}dF(\theta)\Big] \\ &+ \int \gamma(\theta)\Big[\mathbb{I}(s=1)u(c^o(\theta)) + \beta\left(u(c^y(\theta)) - h\left(\frac{y(\theta)}{\theta}\right)\right) + \beta\delta P(\theta)w^1(\theta) + \beta\delta(1-P(\theta))w^0(\theta) - \mathcal{V}(\theta)\Big]d\theta \\ &- \int \Big\{\mu'(\theta)\mathcal{V}(\theta) + \mu(\theta)\beta\Big[h'\left(\frac{y(\theta)}{\theta}\right)\frac{y(\theta)}{\theta^2} + \delta P'(\theta)\left(w^1(\theta) - w^0(\theta)\right)\Big]\Big\}d\theta.\end{aligned}$$

First order conditions include:

$$\left(\mathbb{I}(s=1)u'(c^o(\theta)) - \hat{\lambda}\mathbb{I}(s=1)\right)f(\theta) - \phi\mathbb{I}(s=1)u'(c^o(\theta))f(\theta) + \gamma(\theta)\mathbb{I}(s=1)u'(c^o(\theta)) = 0, \tag{1.85}$$

$$\left(\hat{\beta}u'(c^y(\theta)) - \hat{\lambda}\right)f(\theta) - \phi\beta u'(c^y(\theta))f(\theta) + \gamma(\theta)\beta u'(c^y(\theta)) = 0, \tag{1.86}$$

$$\hat{\beta}P(\theta)J_1(w^1(\theta),1)f(\theta) - \phi\beta P(\theta)f(\theta) + \gamma(\theta)\beta P(\theta) - \mu(\theta)\beta P'(\theta) = 0, \tag{1.87}$$

$$\hat{\beta}(1-P(\theta))J_1(w^0(\theta),0)f(\theta) - \phi\beta(1-P(\theta))f(\theta) + \gamma(\theta)\beta(1-P(\theta)) + \mu(\theta)\beta P'(\theta) = 0, \tag{1.88}$$

and

$$-\gamma(\theta) - \mu'(\theta) = 0. \tag{1.89}$$

At $s=1$ (1.85) and (1.86) give

$$\frac{u'(c^o(\theta))}{\beta u'(c^y(\theta))} = \frac{\hat{\beta}f(\theta) - \phi f(\theta) + \gamma(\theta)}{f(\theta) - \phi f(\theta) + \gamma(\theta)}.$$

Using the definition of τ^i into the previous expression yields

$$\tau^i(\theta) = -\left(\frac{\hat{\beta}}{\beta} - 1\right)\frac{f(\theta)}{f(\theta) - \phi f(\theta) + \gamma(\theta)}.$$

Applying (1.85) into the right hand side gives (1.51).

To obtain the expression for the expected bequest wedge, first note that (1.88) can be written as

$$J_1(w^0(\theta),0)f(\theta) = \phi\frac{\beta}{\hat{\beta}}f(\theta) - \gamma(\theta)\frac{\beta}{\hat{\beta}} - \mu(\theta)\frac{\beta}{\hat{\beta}}\frac{P'(\theta)}{1-P(\theta)}. \tag{1.90}$$

Also, by equation (1.86)

$$\left(1-\frac{\hat{\lambda}}{\hat{\beta}}\right)f(\theta) = \phi\frac{\beta}{\hat{\beta}}f(\theta) - \gamma(\theta)\frac{\beta}{\hat{\beta}},$$

so combining with (1.90) gives

$$J_1(w^0(\theta),0) = \left(1-\frac{\hat{\lambda}}{\hat{u}'(c^y(\theta))}\right) - \frac{\mu(\theta)}{f(\theta)}\frac{\beta}{\hat{\beta}}\frac{P'(\theta)}{1-P(\theta)}. \tag{1.91}$$

Now note that equation (1.86) can be rearranged as

$$\phi = \frac{\hat{\beta}}{\beta} - \frac{\hat{\lambda}}{\beta}\int\frac{1}{u'(c^y(\theta))}f(\theta)d\theta, \tag{1.92}$$

where I used that $\int\gamma(\theta)d\theta = 0$, which follows from (1.89) and $\mu(\underline{\theta}) = \mu(\bar{\theta}) = 0$.

Using the envelope condition

$$J_1(w,s) = \phi \tag{1.93}$$

into the left hand side of (1.92) gives

$$J_1(w^0(\theta),0) = \frac{\hat{\beta}}{\beta} - \frac{\hat{\lambda}}{\beta}\int_{\theta'}\frac{1}{u'(c^y(\theta',w^0(\theta),0))}dF(\theta'). \tag{1.94}$$

The expression for the optimal expected bequest wedge follows by equating (1.91) and (1.94) and applying the definition of $\bar{\tau}^b$.

The derivation of the optimal expected capital tax is very similar. Equations (1.87) and (1.86) can be rearranged as

$$\frac{\hat{\beta}}{\beta}J_1(w^1(\theta),1)f(\theta) = \frac{\hat{\beta}}{\beta}\left(1-\frac{\hat{\lambda}}{\hat{\beta}u'(c^y(\theta))}\right)f(\theta) + \mu(\theta)\frac{P'(\theta)}{P(\theta)}. \tag{1.95}$$

Equation (1.85) and $\int\gamma(\theta)d\theta = 0$ produces

$$\phi = 1 - \hat{\lambda}\int\frac{1}{u'(c^o(\theta))dF(\theta)}.$$

Applying the envelope condition (1.93):

$$J_1(w^1(\theta),1) = 1 - \hat{\lambda}\int_{\theta'}\frac{1}{u'(c^o(w^1(\theta),1,\theta'))}dF(\theta'). \tag{1.96}$$

Combining (1.95) and (1.96) and applying the definition of $\bar{\tau}^k$ gives (1.49).

1.7.2 ADDITIONAL DETAILS ON THE ESTIMATION

DERIVATION OF THE PROBABILITY OF SURVIVAL $\tilde{P}_t(\nu)$

This section describes the derivation of the probability of survival $\tilde{P}_t(\nu)$ in equation (1.57). Suppose time is continuous. Let $\Lambda_t(\nu)$ be the cumulative mortality hazard for an individual of type ν, so that

$$\Lambda_t(\nu) = \int_0^t \eta_s(\nu) ds = \nu \Lambda_t,$$

where $\Lambda_t \equiv \int_0^t \eta_s ds$ is the cumulative mortality hazard for the standard individual.

Under (1.54) and (1.55) it is straightforward to show that

$$\Lambda_t = \frac{1}{\phi} \left(\exp(\phi t) - 1 \right)$$

Equation (1.57) follows by using the previous expression and that $\Lambda_t(\nu) = -\frac{d \log \tilde{P}_t(\nu,\phi)}{dt}$.

CALIBRATION OF ρ

To calibrate ρ, I first divide the wage distribution into quartiles. Following Mankiw et al. (2009) I set the maximum possible wage to \$500.51, but the calibrated value of ρ is not sensitive to this upper bound. I assume that permanent (pre-tax) income is monotonic in hourly wages. This assumption implies that the ranking of individuals across skills is the same as the ranking across permanent income. As a consequence, individuals who belong to a certain quartile of the wage distribution also belong to the corresponding quartile of the permanent income distribution.[47] Within each quartile, I uniformly draw $N_q = 1,000$ individuals, where each individual corresponds to a wage-age pair. For each individual, I then draw a mortality type ν from the log-normal pdf $f(\log(\nu)|\log(\theta))$ over a grid of values for ρ. Then I simulate their life spans using the probability of survival P_t estimated previously. I repeat this procedure $N_s = 100$ times and compute two-year mortality rates by taking averages across simulations. The calibrated value of ρ minimizes the distance between these simulated moments and its counterparts in the data. Figure 1.5 shows the distance between data and simulated moments over a grid of values for ρ.

1.7.3 COMPUTATIONAL APPENDIX

The component planning problem is solved numerically in MATLAB. The optimal control problem embedded in each step of the value function iteration algorithm is solved using GPOPS-II software.

[47] This claim can be easily verified as follows. Let H denote the cumulative distribution function of permanent income $y(\theta)$. If $y'(\theta) > 0$, then $H(y(\theta)) = F(\theta)$. Hence, there is a one-to-one correspondence between the quartiles of the skill distribution and the quartiles of the permanent income distribution.

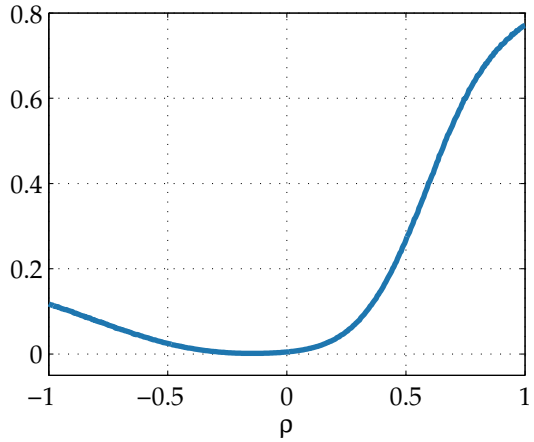


Figure 1.5: Objective Function for Calibration of ρ.

Figure 1.6 plots the computed value functions at the solution for the case with $\hat{\beta} = 1.35$; the other scenario is similar. Note that value functions have a single peak. This property is shown formally in Farhi and Werning (2007) in an environment with one-period lived agents.

To approximate the steady state distribution, I start from an arbitrary value of (w, s) and simulate the model for 10,000 periods. For this step, I interpolate policy functions using thin-plate splines. Figure 1.7 shows the histograms for the simulated time-series along with a fitted Kernel distribution, after dropping the first 500 observations. Notably, there is no mass on neither the upper nor on the lower bounds, so that Ψ^* does not depend on these arbitrary values of the grid. Also, note that the mode occurs around the peaks of the value functions.

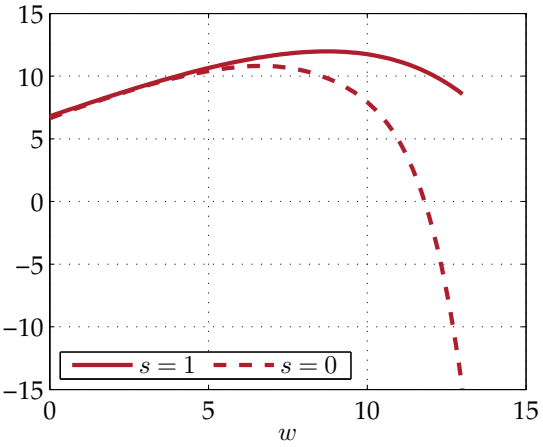


Figure 1.6: Value Functions $J(w, s)$.

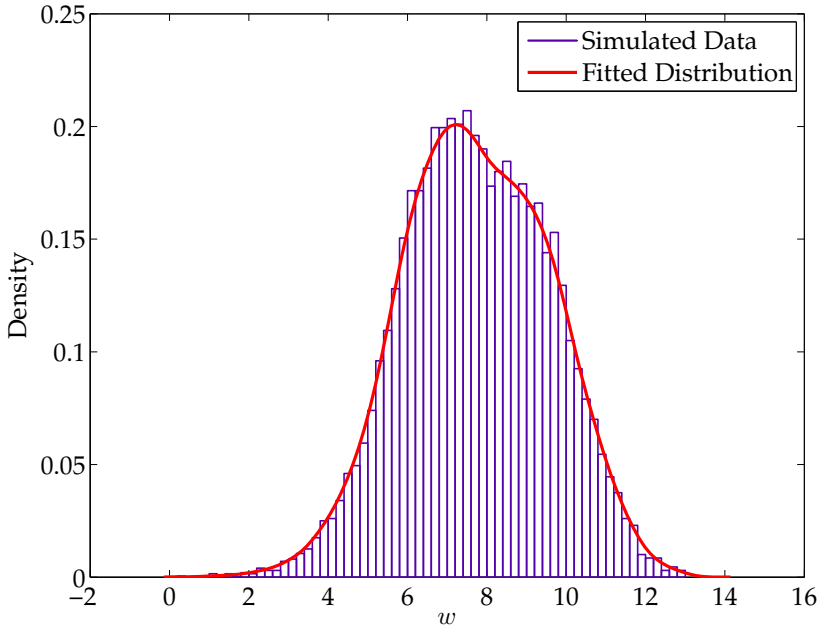


Figure 1.7: Steady State Distribution Ψ^*.

1.7.4 IMPLEMENTATION WITH ANNUITIES

In this section I construct a decentralization which allows households to trade annuities. The purpose is to show that the optimal wealth taxes derived in the body of the paper are robust to the inclusion of annuities in the decentralization.

The annuity market works as follows. In the first period, households purchase annuities in the amount $z_2 \geq 0$. In the second period, individuals receive z_2 units of the consumption good before taxes contingent on survival. Annuity contracts are non-exclusive and linear, with $q \geq 0$ denoting the per unit price of annuities in terms of the consumption good. These contracts are supplied by a continuum of insurers who are Bertrand competitors.[48]

The problem of a dynasty of type θ is

$$\max_{\{c,y_1,k_2,g_2,z_2\}} \quad u(c_1^y) - h\left(\frac{y_1}{\theta}\right) + P(\theta)\delta\Big(u(c_2^o) + \beta u(c_2^y)\Big) + (1 - P(\theta))\delta\beta u(\tilde{c}_2^y) \tag{1.97}$$

subject to

$$\begin{aligned} c_1^y + k_2 + qz_2 &= y_1 - T_1(y_1) + W_1, \\ c_2^o + g_2 &= Rk_2(1 - t_2^k(y_1)) + z_2(1 - t_2^z(y_1)) + SS_2(y_1), \\ c_2^y &= g_2(1 - t_2^i(y_1)), \\ \tilde{c}_2^y &= Rk_2(1 - t_2^b(y_1)) + \tilde{SS}_2(y_1), \end{aligned}$$

[48]The characterization of the annuity market follows Hosseini (2014). Please refer to that study for a detailed justification on the assumptions of non-exclusivity and linear pricing.

where $t_2^z(y_1) : [0, \bar{y}] \to \mathbb{R}$ is a nonlinear tax on annuity returns.

A *competitive equilibrium with annuities* is an allocation $\{c(\theta), y_1(\theta)\}_{\theta\in\Theta}$, a sequence of capital, inter-vivos gifts, and annuity purchases $\{k_2(\theta), g_2(\theta), z_2(\theta)\}_{\theta\in\Theta}$, a tax system $\{T_1, t_2^k, t_2^b, t_2^i, t_2^z, SS_2, \tilde{S}S_2\}$, and an annuity price q^* such that:

1. Each θ-dynasty solves problem (1.97).
2. The government's budget constraint is balanced in every period.
3. Goods and asset markets clear.
4. Annuity insurers make zero profits.

The next proposition shows that it is possible to construct taxes on annuity returns such that optimal taxes on capital, bequests, and inter-vivos transfers are *not affected* by the presence of the annuity market in the decentralization.

Proposition 1.7.1. *Let $\{c^*, y_1^*\}$ solve the relaxed planning problem and satisfy (1.11). There exists an annuity tax schedule such that $\{c^*, y_1^*\}$ can be implemented as a competitive equilibrium with annuities, where for all θ optimal taxes on capital, bequests, and inter-vivos transfers satisfy*

$$t_2^{w*}(y_1^*(\theta)) = \tau^{w*}(\theta), \quad \textit{for} \quad w = \{k, b, i\},$$

with $\{\tau^{w*}(\theta)\}_{w=k,b,i}$ given by (1.22)-(1.24).

Proof. The proof follows by construction. Let $Y_1^* \equiv \{y_1 \in [0, \bar{y}_1] : \exists\theta \text{ such that } y_1 = y_1^*(\theta)\}$, and construct capital, bequest, and inter-vivos taxes as in Proposition 1.2.1. Let the equilibrium price on annuities be:

$$q^* = \frac{P(\bar{\theta})}{R}, \tag{1.98}$$

and define annuity taxes as

$$1 - t_2^{z*}(y_1) = \begin{cases} \frac{q^* u'(\hat{c}_1^y(y_1)}{P(\bar{\theta})\delta u'(\hat{c}_2^o(y_1))}, & \text{if } y_1 \in Y_1^*, \\ 0, & \text{otherwise,} \end{cases} \tag{1.99}$$

where functions $\hat{c}_1^y : Y_1^* \to \mathbb{R}_+$ and $\hat{c}_2^o : Y_1^* \to \mathbb{R}_+$ are defined in Proposition 1.2.1.

I consider a decentralization in which all agents are induced to save $K_2 > 0$ in per capita terms, and make inter-vivos transfers in the amount $G_2 > 0$. As for the annuity demand schedule, let $\bar{z} > 0$ and define:

$$z_2(y_1) = \begin{cases} \bar{z}, & \text{if } y_1 = y_1^*(\bar{\theta}), \\ 0, & \text{otherwise,} \end{cases} \tag{1.100}$$

Finally, construct transfers T_1, SS_2, and $\tilde{SS}_2$ such that the households' budget constraints hold with equality given the previous choices of wealth taxes, capital holdings, inter-vivos gifts, and annuity purchases.

I claim that this construction constitutes a competitive equilibrium with annuities. First, consider the subproblem of a θ-dynasty consisting on choosing consumption allocations for a given level of effective effort y_1. The solution to this problem is characterized by the Euler equations:

$$u'(c_1^y) = \delta R\Big[P(\theta)u'(c_2^o)(1-t_2^{k*}(y_1)) + (1-P(\theta))\beta u'(\tilde{c}_2^y)(1-t_2^{b*}(y_1))\Big],$$

$$u'(c_2^o) = \beta u'(c_2^y)(1-t_2^{i*}(y_1)),$$

and

$$q^* u'(c_1^y) \geq P(\theta)\delta u'(c_2^o)(1-t_2^{z*}(y_1)), \quad \text{with "=" if } z_2 > 0, \tag{1.101}$$

together with the budget constraints of problem (1.97).

Given my construction of wealth taxes, transfers, inter-vivos gifts, and asset purchases, the Euler equations and budget constraints hold at $\{\hat{c}_1^y(y_1), \hat{c}_2^o(y_1), \hat{c}_2^y(y_1), \hat{\tilde{c}}_2^y(y_1), K_2, G_2, z_2(y_1)\}$ for all $y_1 \in Y_1^*$.[49] Incentive compatibility induces each θ-type to choose $y_1^*(\theta)$, and large penalties preclude agents from choosing $y_1^* \notin Y_1^*$. Consequently, optimal consumption and effective effort allocations solve the household's problem.

Notice that the choice of the annuity price in (1.98) and the fact that only individuals with $\theta = \bar{\theta}$ buy annuities, imply that annuity insurers make zero profits. Additionally, all markets clear and the government's budget constraints hold by construction. This completes the proof. □

It is worth noting that under the decentralization of Proposition 1.7.1, annuity markets are not shut down. Instead, individuals at the top of the skill distribution do buy annuities in equilibrium. Moreover, the optimal tax system is such that annuity and capital taxes coincide, so that there is no differential asset taxation in the second period once annuity markets are opened.

[49] Functions $\hat{c}_2^y : Y_1^* \to \mathbb{R}_+$ and $\hat{\tilde{c}}_2^o : Y_1^* \to \mathbb{R}_+$ are defined in Proposition 1.2.1.

Chapter 2

Taxing Atlas: Using Firm Data to Derive Optimal Income Tax Rates

joint with Laurence Ales and Jessie J. Wang

2.1 INTRODUCTION

Income taxation of top income earners is a controversial and recurrent topic in the tax policy debate.[1] However, the large literature in public finance is far from reaching a consensus on what the top tax rate should be. The range of proposed tax rates for top income earners is surprisingly large, ranging from 0% (much lower than the current rate) to 80% (much higher than the current rate).[2] This lack of consensus is in part attributed to the lack of agreement on the magnitude of behavioral tax responses of workers and on the prevalence of highly talented individuals in the population. The main result of this paper is that even with fairly inelastic workers, optimal marginal tax rates for high income earners are in line, if not slightly lower, with what we see in the US today.[3]

In this paper, we model top income earners as managers.[4] Managers are heterogeneous in their skill level, and they operate a span of control technology similar to Rosen (1982). In addition they exert effort, which combined with hired labor, generates output. The usual approach in public finance is to have skills as a productivity parameter for hours worked: more skilled individuals

[1] This is partially due to heightened concerns over the increasing inequality of the income distribution. Indeed the share of income going to the top 1% increased from 9% in 1970 to 23.5% in 2007 (Diamond and Saez (2011)). From Piketty and Saez (2003), the top 1% accounted for 59.8% of average growth in income compared to just 9% of average growth accounted for by the bottom 90% over this period.

[2] Mirrlees (1971) and Sadka (1976) prescribe a zero marginal tax rate at the top. Using different assumptions on the skill distribution, Saez (2001), or Diamond (1998) call for for asymptotic top rates as high as 80%. See Diamond and Saez (2011) for a recent survey.

[3] Saez et al. (2012) report a top 1% marginal rate of approximately 42.5% for 2009. In the Current Population Survey in the same period we find top marginal income tax rates of 33.5% for federal and 5% for state.

[4] Using tax return data Bakija et al. (2012), document that executives, managers, supervisors account for about 40% of the top 0.1% of income earners in recent years. The number grows to 60% including managers and professionals in the financial sector.

can transform hours worked into output at a higher rate. In our environment we maintain this labor productivity channel; however, the skill of the manager will now also determine its managerial ability. This ability will affect the overall productivity of the firm, creating a scale-of-operations effect (see Mayer (1960)). We assume a positive relationship between labor productivity and managerial ability. Modeling skills in this fashion has three important implications. First, more productive managers operate larger firms and receive larger levels of (before-tax) income. In addition, as originally shown by Rosen (1982), the scale-of-operations effect implies that managerial compensation grows at a faster rate than managerial skills. Thus, the distribution of income becomes more positively skewed relative to the distribution of skills. Third, wages of managers (and workers) are endogenous; in particular, they depend on the amount of effort exercised by managers. This is important since any tax aimed at top income earners will impact indirectly their pre-tax wages.

Following the Mirrleesian tradition, we assume that full redistribution is hindered by informational frictions. We consider the case in which the effort and skill level of the manager are private information. On the other hand, firm size and firm output are observable to the policy maker. We characterize the constrained efficient allocation and show that it can be decentralized as a competitive equilibrium with taxes on income and firm size.

Our first result concerns firm size distortions. We provide a formula highlighting how in general the planner will forgo efficiency in the allocation of labor in order to relax incentive constraints and hence reduce informational rents of the managers. We also highlight that these firm level distortions will not always arise. In our benchmark, featuring a positive scale-of-operations effect and an elasticity of substitution between hired labor and managerial effort different than one, they will. We then provide a formula for optimal taxes. This formula will link marginal taxes to primitives of the environment. As in Saez (2001) the formula links marginal rates to the assumed distribution of skills, redistributive motives of the policy maker, and elasticity of labor supply of the manager. In addition, the span of control production function will introduce three novel terms. The first term arises from the fact that the skills and effort of the manager enter asymmetrically into the production function. Given this, the social planner weights the overall impact of the behavioral response of the manager on marginal taxes by how sensitive output is to changes in effort versus changes in skill. The second and third terms arise from the fact that changes in managerial effort impact the marginal product of effort and the marginal product of skills. Hence, the behavioral response of the manager will be dampened by how much his informational rents and wages are impacted by his change in effort.

In order to quantify the effect of the scale-of-operations effect, in turn, to provide an answer to our motivating question on what should taxes be, we need to calibrate the model. There are two items that are important for optimal taxes for which the literature provides little guidance. The first one is the magnitude of the scale-of-operations effect: the importance of managerial ability in determining the overall productivity of the firm. The second term is the distribution of skills. To identify both of these objects we follow the key insights from Rosen (1982). In particular,

we expand his result (adding elastic labor) showing how the scale-of-operations effect can be backed out from the data with knowledge of the elasticity of firm size (in terms of employment) to firm sales and the elasticity of managerial compensation to firm sales. We determine both of these elasticities using Compustat data. The last term concerns the distribution of skills. The usual approach is to invert the distribution of income to uncover the distribution of skills. We take a similar route. However, our environment generates a very nonlinear relationship between skills and income and between skills and firm size. This implies that as the skill of the manager increases by 1%, firm size and managerial income will increase substantially more than that (in a standard Mirrleesian environment, keeping effort fixed, income would increase by only 1%). Given this, using the distribution of firm size to invert the distribution of skills will imply that the tail of the skill distribution is substantially smaller. Assuming a Pareto distribution our estimate on the tail parameter is an order of magnitude larger than what has been previously identified in the literature. As the tail of the distribution of skills gets thinner, taxes at the top decrease.

The optimal top tax rate in our benchmark calibration is 32.4%, which is in the same range (if not slightly lower) than what the current US tax code prescribes. For comparison, absent a scale-of-operations effect the optimal rate would be equal to 65.4%, which exceed the corresponding figure in the US by about twenty percentage points. Additionally, the optimal marginal tax on firm size is positive, progressive, and creates a markup of about 2% over the real wage at the top.

To summarize, this paper makes three contributions. First, we provide normative grounds for why (and when) it is optimal to distort firm sizes in order to reduce informational rents of managers. Second, we compute optimal taxes in a span of control environment as in Rosen (1982) showing that the current US rates are roughly in line with the optimal rates. Third, we illustrate how the determinants of such a tax formula can be calibrated using readily available firm level data as opposed to confidential income data coming from social security or tax return records.

Related literature

This paper touches on two large literatures: the first one concerns managerial compensation and the second one deals with the taxation of managerial income.

The works of Lucas (1978) and Rosen (1982) provide early frameworks where the compensation of the CEO (the owner of the span of control technology) can be analyzed together with the size of the firm. In these models the manager chooses the factors of production to be purchased from the market. Gabaix and Landier (2008), on the other hand, consider a model where the size of the firm is fixed, and the most productive managers are being assigned to the largest firms. What is key in all of these models (for the purpose of optimal taxation) is that they introduce a nonlinear mapping between compensation and skill of the manager. In particular compensation is more skewed than skill. Our contribution with respect this literature is twofold. First, we model the intensive margin of managerial effort (this is a necessary step in order to think about income taxation). Second, we provide a novel calibration strategy that relies on micro evidence from the firm side (Compustat) and from the household side (Current Population Survey).

The literature on optimal taxation of top income earners is vast.[5] Methodologically, our contribution with respect to this literature is that our environment is one in which compensation of the agent (the manager in our case) is endogenous. This is a departure from the classical taxation environment where wages are fixed exogenously. A few exceptions to this assumption are discussed in Stiglitz (1982) where workers of different types interact within an aggregate production function, hence influencing each other wages. In more recent work, Slavík and Yazıcı (2014) focus on the endogenous accumulation of heterogenous forms of capital that interact differently with agents of heterogenous skills. Ales et al. (2014) focus on an assignment problem of workers with heterogenous talents to tasks with heterogenous complexity.

Our approach in this paper is to map top income earners to managers.[6] Given this, Rothschild and Scheuer (2013) and in particular Scheuer (2014) are the papers more related to ours. They consider an environment where agents are characterized by a multidimensional skill/taste vector and decide whether to be a worker or a manager. In the spirit of Stiglitz (1982), a key channel that impacts optimal taxes is the spillover effect between wages of workers and wages of managers. Relative to these papers our emphasis is more quantitative. However there are key differences in the environment that shape optimal taxation. The most important is the role of managerial skill. In our environment, as in these papers, managerial skill impacts how productive the manager is at transforming hours in effective effort that enters in the production function. Differently from the above, though, skill also affects the overall productivity the production function. This in turn increases productivity of the entire firm. We will show that as this second component disappears then optimal taxes revert to the Mirrleesian benchmark as in Scheuer (2014) or Saez (2001). Finally Piketty et al. (2014) consider a model of CEO taxation where, beyond the behavioral labor supply response, the CEO can extract surplus by imposing a negative externality on workers raising his own compensation above his marginal product. In this case, this last channel provides upward pressure on marginal tax rates as the taxes try to correct for this negative CEO externality. In our paper the approach is more in line with the empirical evidence in Kaplan and Rauh (2013) where managers are characterized as having a positive externality on workers rather than a negative one.

The remainder of the paper is organized as follows. In Section 2.2 we describe the environment. In Section 2.3 we characterize Pareto optimality. In Section 2.4 we look at the decentralization of the optimum and derive the optimal tax rate formula. In Section 2.5 we show our identification strategy and calibrate the model. In Section 2.6 we discuss the quantitative results. Section 2.7 looks at additional robustness calculations. In Section 2.8 we analyze optimal firm size distortions, and Section 2.9 concludes.

[5] For a review refer to Mankiw et al. (2009) or Diamond and Saez (2011).

[6] The environment in this paper is be static. For dynamic models that consider the modeling and taxation of entrepreneurial wealth, refer to Quadrini (2000), Cagetti and Nardi (2006), Albanesi (2011) or Shourideh (2012). For a review on the *new dynamic public finance* approach to optimal dynamic taxation refer to Golosov et al. (2007) or Kocherlakota (2010).

2.2 ENVIRONMENT

The economy is static and it is populated by a unit measure of workers and a unit measure of managers. There is a single consumption good. Managers have quasi-linear preferences over consumption c and effort n which is represented by the utility function:

$$U(c,n) = c - v(n),$$

where $v : \mathbb{R}_+ \to \mathbb{R}$ and is twice continuously differentiable with positive derivatives. Workers have preference over consumption and supply labor inelastically. Without loss of generality, we normalize the disutility from effort of the worker to zero and the amount of effective effort supplied to one. Consumption of the worker is denoted by $c^w \in \mathbb{R}_+$.

Managers are heterogeneous with respect to a productivity parameter $\theta \in \Theta$ with $\Theta = \left[\underline{\theta}, \overline{\theta}\right] \subset \mathbb{R}_+ \setminus 0$. Managerial type, θ, is distributed according to the cumulative distribution function $F : \Theta \to [0,1]$ with density function $f : \Theta \to \mathbb{R}_+$. Following Rosen (1982) and Lucas (1978), managers operate a span of control technology. Specifically, managers of type θ hire labor $L(\theta)$ (supplied by workers) and exert managerial effort $n(\theta)$ to produce final output $y(\theta)$ according to:

$$y\big(n(\theta), L(\theta), \theta\big) = \theta^{\gamma} H\left(\theta n\left(\theta\right), L\left(\theta\right)\right), \tag{2.1}$$

where $\gamma > 0$, $H : \mathbb{R}_+ \times \mathbb{R}_+ \to \mathbb{R}_+$ is strictly increasing in both arguments and features continuous derivatives. We assume that H satisfies constant returns to scale.

Managerial skills enter the production function (2.1) in two ways. First, θ is managerial effort-augmenting (since it multiplies $n(\theta)$ within H). Second, θ is total-factor-productivity improving (since $\gamma > 0$). In what follows, we refer to γ as the scale-of-operations effect as in Mayer (1960).[7] This formulation is in line with the one in Rosen (1982) where managers' actions naturally affect the productivity of all workers under their supervision (irrespective of their number). Our technological specification is more general to that in Rosen (1982) given that we also incorporate managerial effort n as an intensive margin. We assume that managerial allocations $c(\theta)$, $y(\theta)$ and $L(\theta)$ are observable, while θ, $n(\theta)$ and $\theta \cdot n(\theta)$ are private information of each agent.

An allocation in this economy is defined as (c^w, c, y, L) where: $c^w \in \mathbb{R}_+$, $c : \Theta \to \mathbb{R}_+$, $L : \Theta \to \mathbb{R}_+$, $y : \Theta \to [0, \bar{y}]$, and $0 < \bar{y} < \infty$. An allocation is feasible if:

$$c^w + \int_{\Theta} c(\theta) dF(\theta) \leq \int_{\Theta} y(\theta) dF(\theta), \tag{2.2}$$

and

$$\int_{\Theta} L(\theta) dF(\theta) \leq 1. \tag{2.3}$$

[7] There is also a large literature connecting the skill of manager and the productivity and size of a firm. See, for example, Bartelsman and Doms (2000) and references therein.

Social welfare is evaluated according to the social welfare function

$$SWF = \Psi(c^w) + \int_\Theta \Psi\left(c(\theta) - v(n(y(\theta), L(\theta), \theta))\right) dF(\theta),$$

where $\Psi : \mathbb{R} \to \mathbb{R}$ is an increasing, differentiable and concave function which summarizes social preferences for redistribution across types. In particular, we will refer to $\Psi'\left(c(\theta) - v(n(y(\theta), L(\theta), \theta))\right)$ as the *social marginal welfare weight* on θ-managers.

2.3 PARETO OPTIMALITY

In this section, we characterize Pareto optimal allocations using a direct mechanism where managers report their types θ to a social planner and are assigned an allocation for consumption $c(\theta)$, output $y(\theta)$ and labor $L(\theta)$ accordingly. Define $n(y(\theta'), L(\theta'), \theta)$ as the level of effort exerted by a manager of type θ who mimics a manager of type θ'. In this case, manager θ will be assigned $L(\theta')$ workers and will be required to produce $y(\theta')$ output. An allocation is *incentive-compatible* when truthful revelation is optimal for all managers, that is:

$$c(\theta) - v(n(y(\theta), L(\theta), \theta)) \geq c(\theta') - v(n(y(\theta'), L(\theta'), \theta)), \quad \forall\, \theta, \theta' \in \Theta. \tag{2.4}$$

Pareto optimal allocations solve the following social planner's problem:

$$\max_{c^w, \{c(\theta), y(\theta), L(\theta)\}_{\theta \in \Theta}} \Psi(c^w) + \int_\Theta \Psi\left(c(\theta) - v(n(y(\theta), L(\theta), \theta))\right) dF(\theta), \tag{PO}$$

$$\text{s.t.} \qquad (2.2), (2.3) \text{ and } (2.4).$$

The social planner's problem (PO) is intractable due to the double infinity of incentive compatibility constraints embedded in (2.4). In this paper we will work with a relaxed version of the planner problem by applying the *first order approach*: we replace the original set of constraints (2.4) with local first order conditions. The validity of the first order approach can be verified ex-post using simulations.[8]

Let $n(y, L, \theta)$ be the effort required by a manager of type θ to generate output y when the number of hired workers is L. Denote by n_y, n_L and n_θ, the first derivatives of n with respect to its first, second and third arguments (with similar notation for its second derivatives). Define $n(\theta) \equiv n(y(\theta), L(\theta), \theta)$ and $U(\theta) \equiv c(\theta) - v(n(\theta))$ for all θ. The relaxed version of the planner's problem is:

$$\max_{c^w, \{U(y(\theta), L(\theta)\}_{\theta \in \Theta}} \Psi(c^w) + \int_\Theta \Psi\left(U(\theta)\right) dF(\theta), \tag{PO-FOC}$$

$$\text{s.t.} \qquad U'(\theta) = -v'(n(y(\theta), L(\theta), \theta)) n_\theta(y(\theta), L(\theta), \theta), \ \forall \theta, \tag{2.5}$$

[8] This approach is fairly standard in the dynamic public finance literature. See Kapička (2013b), Golosov et al. (2013), or Farhi and Werning (2013b). See also Pavan et al. (2014).

$$\int_{\Theta} \left[y(\theta) - c^w - U(\theta) - v(n(y(\theta), L(\theta), \theta)) \right] dF(\theta) = 0, \tag{2.6}$$

$$\int_{\Theta} L(\theta) dF(\theta) = 1. \tag{2.7}$$

The next proposition provides a sufficient condition which guarantees that the first order condition (2.5) implies the incentive compatible condition in (2.4). In that case, since it is trivial to show that (2.2) and (2.3) hold with equality, any solution to (PO-FOC) will also be a solution to (PO).

Proposition 2.3.1. *Suppose that for all $\theta \in \Theta$ the following holds*

$$\frac{v''(n(\theta))}{v'(n(\theta))^2} c'(\theta) + \frac{n_{\theta y}(\theta)}{n_\theta(\theta)} y'(\theta) + \frac{n_{\theta L}(\theta)}{n_\theta(\theta)} L'(\theta) \geq 0. \tag{2.8}$$

Then (2.5) *implies* (2.4).

Proof. See Appendix 2.10.1. □

Relative to the benchmark Mirrleesian environment the sufficient condition in (2.8) also imposes a condition for the allocation of L. The condition will hold, for example, as long as c, y and L are increasing in θ and $n_{\theta L}$ is small enough.[9] From here onwards we assume that (2.8) holds.

Before presenting the decentralization and showing the properties of the optimal tax system we discuss the incentive constraint (2.5) further. The standard Mirrleesian environment features the following relationship between output, type and effort: $y(\theta) = \theta \cdot n(\theta)$. In this case effort required by θ to generate output y is $n(y, \theta) = y/\theta$. So that the term n_θ appearing in the right hand side of (2.5) is given by $n_\theta(y, \theta) = -y/\theta^2 = -n(\theta)/\theta$. In this case, the incentive constraint only depends on the level of effort and does not depend separately on either output or employment. Instead, our environment features a (potential) nonlinear relationship between between output, type, and effort. Also, it features an additional input to production: L. We next provide additional insights on the incentive constraint for this case. Let $h(x) = H(x, 1)$ for all $x \geq 0$. Since H is constant return to scale we have that by definition $y = \theta^\gamma \cdot L \cdot h\left(\frac{\theta n(y,L,\theta)}{L}\right)$, so that $n(y, L, \theta) = h^{-1}\left(\frac{y}{\theta L}\right)\frac{L}{\theta}$. In this case we have

$$n_\theta(y, L, \theta) = -h^{-1}\left(\frac{y}{\theta^\gamma L}\right)\frac{L}{\theta^2} - \frac{L}{\theta}\frac{1}{h'\left(h^{-1}\left(\frac{y}{\theta^\gamma L}\right)\right)}\frac{y}{L}\frac{\gamma}{\theta^{\gamma+1}},$$

which simplifies to

$$n_\theta(y, L, \theta) = -\frac{n(\theta)}{\theta} - \frac{\gamma}{\theta^2}\frac{H(\theta n, L)}{h'\left(\frac{\theta n}{L}\right)}. \tag{2.9}$$

The first term in (2.9) is the same term appearing in the Mirrleesian environment discussed above. Indeed, in the case in which $\gamma = 0$ it is the only term appearing. Hence, a modification of the

[9] It is straightforward to verify that $n_\theta, n_{\theta y} \leq 0$ and $n_{\theta L} \geq 0$.

benchmark Mirrleesian environment featuring a nonlinear production function featuring L with $\gamma = 0$, would leave the incentive constraint unchanged. The second term in (2.9) is novel in this environment. A notable feature is that in general n_θ and hence the right hand side of the incentive constraint (2.5) depends explicitly on the allocation of labor L. This implies that a particular choice of L will have the effect of tightening or relaxing the incentive constraint. This of course at the cost of production efficiency. The distorted choice of L in the sections below will be implemented with a nonlinear tax on firm size (identified with the size of its labor force).

The ability of the planner to affect incentives by changing L will depend on the particular functional form H. Suppose that H is Cobb-Douglas so that: $y = \theta^\gamma(\theta \cdot n)^\alpha L^{1-\alpha}$ for some $0 < \alpha < 1$. We have:

$$n_\theta(y, L, \theta) = -\frac{n(\theta)}{\theta} - \frac{\gamma}{\theta^2}\frac{(\theta \cdot n)^\alpha L^{1-\alpha}}{\alpha(\theta \cdot n)^{\alpha-1}L^{1-\alpha}} = -\frac{n(\theta)}{\theta}(1 + \gamma/\alpha),$$

so that in this case also n_θ assumes the same form as the standard Mirrleesian case. Below we will show that the two cases discussed with n_θ independent of L will provide similar tax recommendation as the standard Mirrleesian case. We will instead focus on the a more general CES formulation for H. This production function differently than the Cobb-Douglas case will feature n_θ depending directly on L and will introduce a motive for the planner to distort the allocation of labor.

2.4 OPTIMAL TAXATION

To characterize optimal income taxes in our framework we introduce a decentralization that relies on nonlinear taxes on firm size (T_L) and nonlinear taxes on income (T). We then determine properties of these tax function that will implement the allocation originating from (PO-FOC).

In the decentralized environment managers of type θ solve the following problem taking wages and tax rates as given:

$$U(\theta) = \max_{c,y,L} \; c - v(n(y, L, \theta)) \tag{MP}$$

$$\text{s.t.} \qquad c \leq y - wL - T_L(wL) - T(y - wL - T_L(wL)), \tag{2.10}$$

where $w \in \mathbb{R}_+$ is the real wage, $T : \mathbb{R}_+ \to \mathbb{R}$ is a nonlinear income tax and $T_L : \mathbb{R}_+ \to \mathbb{R}$ is a nonlinear tax on firm size.[10] Since workers in our environment supply labor inelastically their problem is characterized by a simple budget constraint $c^w = w - \phi$, where ϕ is a transfer to the worker. We can now define a competitive equilibrium for our environment.

Definition 2.4.1. *For a given level of government consumption G, a tax distorted competitive equilibrium is an allocation $\{c, y, L\}$, a wage w, and a tax system $\{T, T_L, \phi\}$ such that:*

1. *Taking as given $\{w, T, T_L\}$ each θ-manager solves (MP);*

[10] As we will show next, the optimum features firm size distortions which translates into $T_L' \neq 0$. This explains the necessity of such fiscal instrument for our implementation.

2. *Worker's budget constraint holds:* $c^w = w - \phi$;

3. *Goods and labor markets clear: equations* (2.6) *and* (2.7) *hold;*

4. *Government's budget constraint is balanced:*

$$\int \left[T\left(y(\theta) - wL(\theta) - T_L(wL(\theta))\right) + T_L(wL(\theta))\right] dF(\theta) = G. \tag{2.11}$$

By applying a version of the *taxation principle* we can rewrite the planner's problem as one in which tax functions are chosen directly, subject to the constraint that they induce a competitive equilibrium (see Guesnerie (1981)). Imposing $G = 0$, the resulting problem is:[11]

$$\max_{c^w, \{T(\cdot), T_L(\cdot)\}_{\theta\in\Theta}} \Psi(c^w) + \int_\Theta \Psi\left(U(\theta)\right) dF(\theta), \tag{DEC}$$

$$\text{s.t.} \qquad (2.11),$$

$$c^w = w - \phi, \tag{2.12}$$

$$\int L(\theta) dF(\theta) = 1, \tag{2.13}$$

$$U(\theta) \text{ solves } (MP), \quad \forall\, \theta \in \Theta. \tag{2.14}$$

In what follows, we analyze optimal taxes in our framework.

2.4.1 FIRM SIZE TAXATION

We begin by looking at the distortions on firm size implied by the constrained efficient allocation. Such distortions provide a normative rationale for firm size taxation in the decentralization. Assuming differentiability of T and T_L, first order conditions from the manager's problem (MP) give:

$$w(1 + T_L'(wL(\theta))) = y_L(\theta n(\theta), L(\theta), \theta), \tag{2.15}$$

where y_L is the marginal product of the worker.[12]

Equation (2.15) shows that if $T_L'(wL(\theta)) \neq 0$ for some θ, the worker's marginal product will not be equalized across firms, implying the break down of the well known Diamond and Mirrlees (1971) productive efficiency result. The reason behind the willingness of the planner to distort labor allocations lies in the ability of relaxing incentive constraints by affecting $L(\theta)$. To see this point, let $n(\theta, \theta')$ be the effort of a manager of type θ misreporting being type θ'. Since firm output $y(\cdot)$ and employment $L(\cdot)$ are observable, $n(\theta, \theta')$ must satisfy the following relationship:

$$(\theta')^\gamma H(\theta' \cdot n(\theta'), L(\theta')) = \theta^\gamma H(\theta \cdot n(\theta, \theta'), L(\theta')),$$

[11] Assuming $G = 0$ is without loss of generality for our analytical results.
[12] Here we use that $y_L = -n_L/n_y$, which simply follows from the implicit function theorem.

and since H is homogenous of degree one, we have that

$$n(\theta,\theta') = h^{-1}\left(\left(\frac{\theta'}{\theta}\right)^{\gamma} h\left(\frac{\theta' n(\theta')}{L(\theta')}\right)\right)\cdot\frac{L(\theta')}{\theta}, \tag{2.16}$$

which shows that, in general, the level of $L(\theta')$ impacts $n(\theta,\theta')$ and hence the benefit of a deviation from θ to θ'.

Remark 2.4.1. *There are two important examples in which $n(\theta,\theta')$ is independent of $L(\theta')$. The first case is when the scale-of-operations effect is shut down, i.e. $\gamma = 0$. In this case, $n(\theta,\theta') = \frac{\theta'}{\theta}n(\theta')$ and our environment collapses to the standard Mirrleesian model. A second example emerges when H is Cobb-Douglas. In particular, if $H(\theta\cdot n(\theta), L(\theta)) = (\theta n(\theta))^{\alpha}L(\theta)^{1-\alpha}$, from (2.16) we obtain $n(\theta,\theta') = \left(\frac{\theta'}{\theta}\right)^{\frac{\gamma+\alpha}{\alpha}} n(\theta')$.*

The next proposition provides a formula for optimal firm sizes distortions and gives conditions under which it is optimal not to distort firm level employment.

Proposition 2.4.1. *Let $\{y(\theta), n(\theta), L(\theta)\}_{\theta\in\Theta}$ be a solution of (PO-FOC) and let T_L^* be a solution to (DEC). Then:*

1. *Optimal marginal firm size distortions satisfy:*

$$T_L'^*(wL(\theta)) = \frac{f(\theta) - \mu(\theta)v'(n(\theta))\frac{n_{\theta L}(\theta)}{\lambda^l}}{f(\theta) + \mu(\theta)v'(n(\theta))\frac{n_{\theta y}(\theta)}{\lambda^r}} - 1, \tag{2.17}$$

   where $\mu(\theta) \geq 0$ is the multiplier on the incentive constraint (2.5) and $\lambda^l > 0$ and $\lambda^r > 0$ are the multipliers on (2.7) and (2.6), respectively.

2. *If there is a differentiable $f : \mathbb{R}_+ \to \mathbb{R}_+$ such that $n_\theta = f(n(\theta))$ for all $\theta \in \Theta$, then $T_L'^*(wL(\theta)) = 0$ for all $\theta \in \Theta$.*

Proof. See Appendix 2.10.2. □

Part 1 of Proposition 2.4.1 illustrates how in general a binding incentive constraint (with multiplier $\mu(\theta)$) generates nonzero firm level distortions. Part 2 provides sufficient conditions for distortions to disappear. Following the discussion at the end of Section 2.3 it is immediate that this condition is satisfied in the case in which $\gamma = 0$ or in the case with $H(\cdot,\cdot)$ is Cobb-Douglas.[13]

[13] Scheuer (2014) considers an environment similar to ours with two important differences. First, the firm level production function features $\gamma = 0$. Second, workers and managers are heterogeneous in two dimensions. He considers two cases: one in which the government can tax workers and managers differently, and one in which it cannot. This second case features firm level distortions, but it is not due to the presence of L in the incentive constraint but from the presence of an additional *no-discrimination* constraint absent in our environment.

2.4.2 INCOME TAXATION

We now move to income taxation. For the rest of the analysis, we make the following assumption on preferences:

Assumption 2.4.1. *The disutility for effort is iso-elastic: $v(n) = n^{1+\frac{1}{\varepsilon}}/(1+\frac{1}{\varepsilon})$, where ε is the Frisch elasticity of labor supply.*

Assuming differentiability of T', first order conditions from the manager's problem (MP) are:

$$1 - T'(\pi(\theta)) = v'(n(y(\theta), L(\theta), \theta))n_y(y(\theta), L(\theta), \theta), \tag{2.18}$$

where $\pi(\theta) \equiv y(\theta) - wL(\theta) - T_L(L(\theta))$ corresponds managerial income of type θ. The next proposition characterizes the optimal marginal income tax rates. To simplify notation let $n(\theta) \equiv n(y(\theta), L(\theta), \theta)$ and $y(\theta) \equiv y(n(\theta), L(\theta), \theta)$ with similar notation for the derivatives y_n, y_θ, y_{nn} and $y_{n\theta}$.

Proposition 2.4.2. *Let $\{y(\theta), n(\theta), L(\theta)\}_{\theta\in\Theta}$ be a solution of (PO-FOC). Let $\{T', T_L^*, w\}$ be a solution to (DEC). We have that for all θ:*

$$\frac{T'(\pi^*(\theta))}{1 - T'(\pi^*(\theta))} = \frac{1 - F(\theta)}{\theta f(\theta)} \cdot \left(1 - \frac{D(\theta)}{D(\underline{\theta})}\right) \cdot A(\theta) \cdot \left(\frac{1}{\varepsilon} + B(\theta)\right) \tag{2.19}$$

where

$$A(\theta) \equiv \frac{y_\theta(\theta)}{y_n(\theta)} \frac{\theta}{n(\theta)}, \qquad B(\theta) \equiv \left[\frac{y_{\theta n}(\theta)}{y_\theta(\theta)} - \frac{y_{nn}(\theta)}{y_n(\theta)}\right] n(\theta),$$

and

$$D(\theta) \equiv \frac{1}{1 - F(\theta)} \int_\theta^\infty \Psi'\left(U^*(\theta)\right) dF(\theta),$$

where $\pi^*(\theta) \equiv y(\theta) - wL(\theta) - T_L^*(L(\theta))$.

Proof. See Appendix 2.10.2. □

Equation (2.19) reveals the main forces behind optimal marginal income taxes in our framework. The first two terms are well known and are present in the seminal contribution of Saez (2001) expressing the optimal taxation results of Mirrlees (1971). The first term refers to the effect of shape of the skill distribution on marginal tax rates. In particular, high marginal taxes at θ are attractive as the mass of managers above θ, $(1 - F(\theta))$, is large; but the resulting distortion is proportional to the mass of individuals at θ and to their productivity level, explaining the negative dependence on $\theta f(\theta)$. The second term summarizes the impact on marginal taxes of the redistributive tastes of the government, which are embedded into $D(\theta)$.

The remaining term in (2.19), $A(\theta)\left(1/\varepsilon + B(\theta)\right)$, represents the impact on taxes arising from the behavioral response of managers. The term $1/\varepsilon$ represents the standard labor response of agents to changes in after tax wages. Clearly, a high elasticity (embodied in a high value of ε) will

translate into a large response of agents to taxation (it is easy to see that $A(\theta) > 0$), and hence will lower the optimal marginal tax rates. This term and this logic is present in the environment analyzed in Saez (2001). The novelty in our environment is the presence of the terms summarized by $A(\theta)$ and $B(\theta)$. These terms embody the interaction between skills and managerial effort through the span of control technology described in the previous section. We discuss each of these terms in turn.

The term $A(\theta)$ is the ratio of two elasticities: the output elasticity of skill over the output elasticity of effort. When $A(\theta)$ is less than one, output is more responsive to a one percent change effort than to a one percent difference in skills. Lower is the value of $A(\theta)$ lower will be marginal tax rates. This is because any changes in the behavioral responses of manager θ will have a relatively large impact on the output generated by θ. At the opposite end, a high value for $A(\theta)$ will imply that the marginal contribution of the manager of type θ to output arises more from his type than from his own effort. In this case, since output is relatively unaffected by changes in effort, high taxes will be warranted.

The term $B(\theta)$ is the difference between the elasticity of the marginal product of θ with respect to managerial effort: $y_{\theta n}(\theta)/y_\theta(\theta)$ minus the elasticity of the marginal product of n also with respect to managerial effort: $y_{nn}(\theta)/y_n(\theta)$. To understand this term is helpful to think about the incentives of a manager in the decentralized environment considered. Let's consider the first term $y_{\theta n}(\theta)/y_\theta(\theta)$. As this elasticity grows, the larger is the impact on the marginal value of managerial skill to changes in the manager's effort. A large value of this elasticity dissuades the manager to reduce his effort. Given this, marginal taxes induce a smaller behavioral effect and higher marginal tax rates are warranted. Finally, a large absolute value of the elasticity $-y_{nn}(\theta)y_n(\theta)$ (recall that $y_n n < 0$) implies that given changes to effort there will be large changes in the marginal product. This implies that as effort decreases, the marginal product of labor increases so that the distortionary effect of marginal taxes are dampened. This effect, as the previous one, will be a force for higher marginal taxes.

It is illustrative to compare optimal taxes prescribed by (2.19) to a known benchmark. In Mirrlees (1971), the technology is given by $y = \theta \cdot n$ so that $A(\theta) = B(\theta) = 1$. In this case we obtain the classical tax formula from Diamond (1998) or Saez (2001):

$$\frac{T'(\pi^*(\theta))}{1 - T'(\pi^*(\theta))} = \frac{1 - F(\theta)}{\theta f(\theta)} \left(1 - \frac{D(\theta)}{D(\underline{\theta})}\right) \left(\frac{1}{\varepsilon} + 1\right). \tag{MDS}$$

As the following proposition shows the Mirrlees-Diamond-Saez optimal tax formula (MDS) generalizes to any environment where skills are only effort-augmenting (for example in the case in which $\gamma = 0$ in our environment):

Corollary 2.4.1. *Suppose that managers operate a technology without the scale-of-operations effect ($\gamma = 0$) so that $y(\theta) = H(\theta \cdot n(\theta), L(\theta))$. Then if (2.8) is satisfied, at any Pareto optimum $T'(\pi^*(\theta))$ satisfies (MDS).*

Proof. See Appendix 2.10.2 □

To lay the groundwork for our quantitative analysis we make the following parametric assumption on the production function:

Assumption 2.4.2. *The production function has constant elasticity of substitution:*

$$y\big(n(\theta), L(\theta), \theta\big) = \theta^{\gamma}\left[\beta(\theta n(\theta))^{\rho} + (1-\beta)L(\theta)^{\rho}\right]^{\frac{1}{\rho}},$$

where $\rho \in [-\infty, 1]$ *and the elasticity of substitution between* $\theta n(\theta)$ *and* $L(\theta)$ *is given by* $\sigma = \frac{1}{1-\rho} \in [0, \infty]$.

The following corollary characterizes optimal income taxes under Assumption 2.4.2.

Corollary 2.4.2. *Suppose Assumption 2.4.2 holds and that (2.8) is satisfied. Then at any Pareto optimum,* $T'(\pi^*(\theta))$ *satisfies*

$$\frac{T'(\pi^*(\theta))}{1-T'(\pi^*(\theta))} = \frac{1-F(\theta)}{\theta f(\theta)}\left(1-\frac{D(\theta)}{D(\underline{\theta})}\right)\left(\frac{1}{\varepsilon}+1+\frac{\gamma}{1-\kappa^*(\theta)}\left(\frac{1}{\varepsilon}+1-\rho\kappa^*(\theta)\right)\right), \tag{2.20}$$

where $\kappa^*(\theta) \equiv y_L^*(\theta)L^*(\theta)/y^*(\theta)$ *is the share of labor costs to total sales for managers of type* θ.

Proof. See Appendix 2.10.2. □

In the limit, the tax formula in (2.20) can be simplified further assuming the following:

Assumption 2.4.3.

(a) The skill distribution has a right Pareto tail with parameter $\frac{1}{a} > 0$*:*

$$\lim_{\theta\to\infty}\frac{1-F(\theta)}{\theta f(\theta)} = \frac{1}{a}.$$

(b) There is zero marginal social weight at the top: $\lim_{\theta\to\infty} D(\theta) = 0$.

Taking limits on (2.20) and using Assumption 2.4.3 we get an expression for the optimal marginal tax rate at the top.

Corollary 2.4.3. *Suppose Assumption 2.4.2 and 2.4.3 hold and that (2.42) is satisfied. Then at any Pareto optimum the marginal tax at the top* $T'(\pi^*(\infty))$ *satisfies:*

$$T'(\pi^*(\infty)) = \frac{1}{1+a\left[\frac{1}{\varepsilon}+1+\frac{\gamma}{1-k(\infty)}\left(\frac{1}{\varepsilon}+1-\rho k(\infty)\right)\right]^{-1}}. \tag{2.21}$$

Equation (2.21) will be the key equation we will take to the data in the next section. In Section 2.10.6 we compare the above tax formula with the benchmark case analyzed in Saez (2001). In that section we also emphasize the key role that firm level distortion play for the derivation of our optimal tax formula.

2.5 IDENTIFICATION AND CALIBRATION

The tax formula in (2.21) provides insights on the forces that shape top marginal tax rates. Our next goal is to quantify these forces. In this section we describe how to estimate the parameters required to evaluate equation (2.21). These include: the elasticity of labor supply ε; the degree of substitutability between the two types of labor inputs ρ; the Pareto tail parameter for skill distribution a, and the scale-of-operations parameter γ. The first two will be taken from the literature, while the last two will be backed out directly from the data.

There is a vast literature estimating the elasticity of labor supply.[14] We follow the guidelines in Chetty et al. (2011) and set ε equal to 0.5 for our benchmark calculation. When computing taxes we will also show the effect of a wide range of labor elasticities.

In the production function the elasticity of substitution between worker's effort and managerial effort is given by σ. Although there is a large literature documenting the impact of managerial and executive quality on firm output, there is little evidence on how managerial choices such as effort substitute with hired factors of production such as the number of workers. A possible strategy to identify σ is to use the aggregate behavior of wages over time. For example Katz and Murphy (1992) in an aggregate production function estimate an elasticity of substitution between skilled and unskilled worker equal to 1.4 (see also Acemoğlu and Autor (2011)). Replicating their approach, partitioning the population in workers in managerial positions or not, we arrive at an estimate of the elasticity of substitution between workers and managers approximately equal to 4 (see Section 2.7.1 for details about the data). However this approach fails to recognize which workers are assigned to which manager. In our benchmark we set $\rho = 0.8$, implying $\sigma = \frac{1}{1-\rho} = 5$. In Section 2.5.2 we will demonstrate how the assumption of $\gamma > 0$ implies a restriction on σ being strictly greater than one.[15] Finally in Section 2.6 we also perform robustness for alternative values of σ.

We next develop equilibrium restrictions that link the value of γ to observables. Using these restrictions we determine the value of γ to firm level observable moments. Finally in Section 2.5.3 we estimate the value for a.

2.5.1 FIRM LEVEL ELASTICITIES

In this section we follow Rosen (1982) and derive an equilibrium restriction that relates γ to observables. Relative to the environment specified in Rosen (1982) we include income taxation

[14] For prime age males MaCurdy (1981) and Altonji (1986) estimate an elasticity between 0 to 0.54. Saez (2003) using the NBER tax panel from 1979 to 1981, estimates a labor elasticity of 0.25. Similar ranges are estimated by Blundell et al. (2012) and French (2005). Chetty et al. (2011) find values equal to 0.5 on the intensive margin and 0.25 on the extensive margin. In the macro literature King and Rebelo (1999) and Prescott (2004) find values between 2 to 4.

[15] In particular we argue that in order to be compatible with firm level data, we have to impose $\sigma > 3$.

and an elastic margin for effort.[16] As in Section 2.3 we can write

$$y(\theta) = \theta^{\gamma} \cdot L(\theta) \cdot h\left(\frac{\theta n(\theta)}{L(\theta)}\right), \tag{2.22}$$

where $h\left(\frac{\theta n}{L}\right) \equiv H\left(\frac{\theta n}{L}, 1\right)$, so that $h' > 0$ and $h'' < 0$.

Consider a competitive equilibrium where the manager faces a linear tax τ on her income, pays wage w to each unit of labor input L, and gets a fraction $\chi \in (0,1]$ of total profits.[17] The special case $\chi = 1$ implies that there is no separation between ownership and control, which is the case considered in Rosen (1982). We can write the θ-manager's problem as:

$$\max_{\{L,n,\pi\}} (1-\tau)\chi\pi - v(n) \tag{2.23}$$

$$\text{s.t.} \qquad \pi = \left[\theta^{\gamma} L h\left(\frac{\theta n}{L}\right) - wL\right].$$

The first order conditions with respect to L and n in (2.23) are given by:

$$\theta^{\gamma}\left[h\left(\frac{\theta n}{L}\right) - \frac{\theta n}{L}h'\left(\frac{\theta n}{L}\right)\right] = w, \tag{2.24}$$

and

$$(1-\tau)\chi\theta^{\gamma+1}h'\left(\frac{\theta n}{L}\right) = v'(n). \tag{2.25}$$

Equations (2.24) and (2.25) together imply the following Lemma. Here we relate the behavior of observables such as labor size, output and profits grow as skill of the manager grows. It is worthwhile observing that as long as the share of ownership is constant across θ then the level of ownership (χ) does not impact these growth rates hence it will not bias our estimation strategy.

Lemma 2.5.1. *Let $\{L(\theta), n(\theta), \pi(\theta)\}$ solve the θ-manager's problem in (2.23). Then the following relationships hold:*

$$\frac{d\ln L(\theta)}{d\ln\theta} = 1 + \frac{\gamma\sigma}{1-\kappa(\theta)} + \varepsilon\left(1 + \frac{\gamma}{1-\kappa(\theta)}\right), \tag{2.26}$$

$$\frac{d\ln y(\theta)}{d\ln\theta} = 1 + \gamma + \varepsilon\left(1 + \frac{\gamma}{1-\kappa(\theta)}\right) + \frac{\kappa(\theta)}{1-\kappa(\theta)}\gamma\sigma, \tag{2.27}$$

$$\frac{d\ln \pi(\theta)}{d\ln\theta} = \left(1 + \frac{\gamma}{1-\kappa(\theta)}\right)(1+\varepsilon). \tag{2.28}$$

where $\kappa(\theta) \equiv wL(\theta)/y(\theta)$.

Proof. See Appendix 2.10.3. □

Lemma 2.5.1 reveals some key properties of the scale-of-operations effect that are present in our

[16] In this Section we are implicitly assuming that there are no distortions in data that might affect firm size.

[17] The assumption of managers being subject to a constant marginal tax rate is motivated by the progressivity of the US income tax system together with managers, in general, being located at the top of the income distribution.

environment. Equation (2.26) and (2.27) show that since $\gamma > 0$, the elasticity of firm size (in terms of employment L) and sales (in terms of total output γ) with respect to θ are greater than one. This implies that given a distribution of types the respective distribution of firm size and firm sales will be more skewed. The same observation holds for equation (2.28) relating profits for the manager versus his skill θ. It is easy to see that in the case with inelastic labor $\varepsilon = 0$ then the above three equation map to equations (14)–(16) in Rosen (1982).

The estimation of a will be based on equation (2.26) (and on equation (2.28) in the robustness section). The estimation of γ instead relies on equations (2.26) and (2.27). With these equations we can relate the value of γ to: $\frac{d\ln\pi(\theta)}{d\ln y(\theta)}$ the elasticity of manager compensation to firm output; and the elasticity of firm size with respect to sales $\frac{d\ln L(\theta)}{d\ln y(\theta)}$ which will be determined from the data. To this end we will use the following Corollary which is simply derived from Lemma 2.5.1.

Corollary 2.5.1. *For any solution of the θ-manager's problem in* (2.23) *the following relationships hold:*

$$\frac{d\ln L(\theta)}{d\ln y(\theta)} = \frac{(1-\kappa(\theta))(1+\varepsilon)+\gamma(\sigma+\varepsilon)}{(1-\kappa(\theta))(1+\gamma+\varepsilon)+\gamma(\kappa(\theta)\sigma+\varepsilon)}, \tag{2.29}$$

$$\frac{d\ln \pi(\theta)}{d\ln y(\theta)} = \frac{(1-\kappa(\theta)+\gamma)(1+\varepsilon)}{(1-\kappa(\theta))(1+\gamma+\varepsilon)+\gamma(\kappa(\theta)\sigma+\varepsilon)}. \tag{2.30}$$

From (2.29) and (2.30) we obtain

$$1-\kappa(\theta) = \frac{1-\frac{d\ln L(\theta)}{d\ln y(\theta)}}{\frac{d\ln \pi(\theta)}{d\ln y(\theta)} - \frac{d\ln L(\theta)}{d\ln y(\theta)}}, \tag{2.31}$$

rearranging equation (2.29), we have:

$$\gamma = \frac{\left(1-\frac{d\ln L(\theta)}{d\ln y(\theta)}\right)(1-\kappa(\theta))(1+\varepsilon)}{\frac{d\ln L(\theta)}{d\ln y(\theta)}(1-\kappa(\theta)+\kappa(\theta)\sigma+\varepsilon)-(\sigma+\varepsilon)}. \tag{2.32}$$

Substituting (2.31) into equation (2.32), we obtain an expression for γ as a function of firm-level elasticities, and parameters σ and ε.

Proposition 2.5.1. *For any solution of the θ-manager's problem in* (2.23) *the following relationship holds:*

$$\gamma = \frac{1-\frac{d\ln L(\theta)}{d\ln y(\theta)}}{\frac{d\ln L(\theta)}{d\ln y(\theta)} - \frac{d\ln \pi(\theta)}{d\ln y(\theta)}\frac{\sigma+\varepsilon}{1+\varepsilon}}. \tag{2.33}$$

Equation (2.33) in Proposition 2.5.1 form the basis for the estimation of γ. To evaluate γ, from (2.32) we require the elasticity of firm size with respect to sales $d\ln L(\theta)/d\ln y(\theta)$ and the elasticity of managerial compensation with respect to firm size $d\ln\pi(\theta)/d\ln y(\theta)$.[18]

[18]From equation (2.33) it is clear that given data on firm elasticities, not all parameter combinations of (σ,ε) will give a positive value for γ. In Figure 2.2 we show the admissible values of σ and ε that, given the estimated elasticities, are consistent with a positive value of γ.

2.5.2 ESTIMATING γ

We begin with estimating the elasticity of firm size with respect to sales. We proxy the firm workforce with the number of non managerial employees. Our model of the firm does not feature hierarchies: a single manager controls a single firm. A direct application to data would then imply that the number of non managerial employees is equal to total employment minus one. However when considering larger firms is natural to think that multiple top executives are responsible for the operations of a firm. In this case we need to consider the number of top executives. So that for firm i at time t we consider the following relationship:

$$\ln(\text{Sales}_{i,t}) \approx \alpha_0 + \alpha_1 \ln\left(\text{Employees}_{i,t} - \text{Number of top executives}_{i,t}\right),$$

If the number of top executives is a constant fraction of the total number of employees, then looking at the total number of employees as the sole dependent variable will provide a unbiased estimate of α_1. In our environment the span of control is increasing with managerial skill. In this sense in our environment the single manager is associated, as his skill increases, to a larger firm. Given this we consider a fixed number of top executives across all firms. For large firms the assumption on the exact number for top executives will have small effect on the elasticity estimates. This can be seen easily rearranging the above. We have

$$\begin{aligned}\ln(\text{Sales}_{i,t}) &\approx \alpha_0 + \alpha_1 \ln\left(\text{Employees}_{i,t}\right) + \alpha_1 \ln\left(1 - \frac{\text{Number of top executives}}{\text{Employees}_{i,t}}\right) \approx \\ &\approx \alpha_0 + \alpha_1 \ln\left(\text{Employees}_{i,t}\right).\end{aligned}$$

We look at data from publicly traded firms in COMPUSTAT. The sample is constructed at an annual frequency from 2000 to 2012. The sample is comprised of US publicly traded firms. Data of firm sales is taken from *Gross Sales* in the Income Statement; data on the total number of employees is taken from the *Employees* item. Nominal variables are deflated using the CPI for all urban consumers, all goods. For our sample selection we drop firms that report negative or zero sales and firm with duplicate CUSIP for a given year. For every year we rank all remaining firms by size. In our benchmark calculation we consider firms above (and including) the median size. This is done to ensure that the exact number specified for top executive has little effect on the estimates. As a benchmark we assume that the number of top executives is 20. In Figure 2.1 we also report estimates changing the number of top executives from 1 to 50. Also in the Figure is the comparison between our benchmark estimation and the case where firms below the median size are included as well. We create division dummies based on Standard Industrial Classification (SIC) as defined by the Occupational Safety & Health Administration.[19] In our benchmark specification

[19]Division refer to industry groupings. The 10 divisions considered are: Agriculture, Forestry and Fishing; Mining; Construction; Manufacturing; Transportation , Communications, Electric, Gas and Sanitary Services; Wholesale Trade; Retail Trade; Finance, Insurance and Real Estate; Services; Public administration.

we then consider the following linear relationship:

$$\ln y_t(\theta_i) = \alpha_0 + \alpha_1 \ln L_t(\theta_i) + \sum_{j=1}^{10} \alpha_{3,j} \text{Div}_j + \varepsilon_{i,t}. \tag{2.34}$$

where $\ln y_t(\theta_i)$ is the log of firm sales; $\ln L_t(\theta_i)$ is the log of firm size measure (total number of employees-20) and Div_j are the dummy variables for each division. We estimate a value of $\frac{d \ln y}{d \ln L} = .951$ (0.002), where with $\frac{d \ln y}{d \ln L}$ we denote the average value of $\frac{d \ln y(\theta)}{d \ln L(\theta)}$ in our sample. The estimated elasticity is consistent with the *making-do-with-less* effect which implies a coefficient smaller than one as in Lazear et al. (2013).

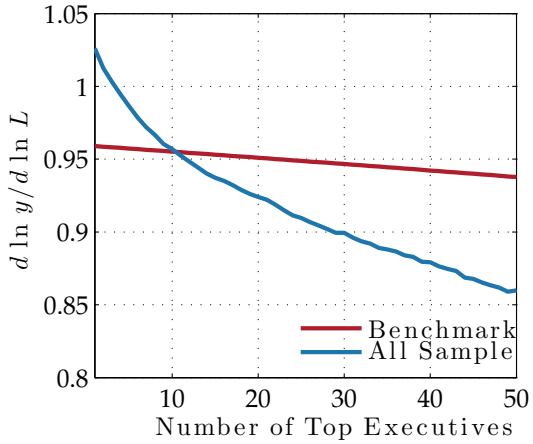


Figure 2.1: Estimates of $\frac{d \ln y}{d \ln L}$ by number of top executives. "Benchmark" refers to estimates of (2.34) using firms above the median size, "All Sample" refers to estimates including firms below the median .

In Table 2.1 we report details about our benchmark estimation and additional robustness checks. Column (1) displays the benchmark. Columns $(2) - (4)$ look at the impact of extending the time period and the effect of either industry or year dummies. Columns $(5) - (6)$ look at the effect of changing the decile of firm size included. We observe that our estimate, with either slightly larger or smaller estimates, is robust to changes in specification.

We next move to estimate the average elasticity of manager compensation to firm output: $\frac{d \ln \pi}{d \ln y}$. Starting from Roberts (1956), there is a vast literature estimating the empirical elasticity of managerial compensation with respect to firm size in the cross-section.[20] The literature has highlighted an empiric al regularity usually denoted as "*Roberts's Law*," which states that on average managerial compensation is proportional to a power of $1/3$ on the own firm size (which could be measured by total firm value, firm sales, etc.). In particular, Gabaix and Landier (2008) estimate that the elasticity of managerial compensation with respect to sales is 0.21 and the elasticity of managerial compensation with respect to firm value is estimated to be equal to 0.34,

[20] Also refer to Lewellen and Huntsman (1970), Baker et al. (1988), and Frydman and Saks (2010).

Table 2.1: Estimating the Elasticity of Firm Size With Respect to Sales

	ln(Sales)					
	(1)	(2)	(3)	(4)	(5)	(6)
ln(Workers)	0.951	0.933	0.956	0.968	0.912	0.972
	[0.002]	[0.003]	[0.001]	[0.001]	[.0008]	[0.004]
Year dummy				Yes	Yes	
Division dummy	Yes		Yes	Yes	Yes	Yes
All time period			Yes	Yes	Yes	
Deciles Included	≥ 5	≥ 5	≥ 5	≥ 5	All	≥ 8
Observations	50,267	50,267	171,044	171,044	265,764	25,131
R^2	0.77	0.71	0.79	0.80	0.84	0.69

Notes: Estimates of α_1 in (2.34). Column (1) displays benchmark calculation using COMPUSTAT data (2000-2012). "Year dummy" denotes the inclusion or not of yearly dummies. "Division dummy" highlights the inclusion or not in (2.34) of dummies based on Standard Industrial Classification (SIC) from Occupational Safety & Health Administration. "All time period" denotes the usage of the entire dataset up to 1950. "Decile Included" denote the sample of firms by size included in the estimation of α_1. We report standard errors in square parentheses.

close to the *Roberts's Law*.[21] In our benchmark calculation we set $\frac{d\ln\pi}{d\ln y} = 0.34$. In Section 2.7.1 we present an alternative calibration strategy that instead does not require computing $\frac{d\ln\pi}{d\ln y}$ but will rely on aggregate data on the relative compensation share of managers versus workers.

We can now determine γ using equation (2.32) in Proposition 2.5.1. we get:

$$\gamma = \frac{1 - \frac{1}{0.951}}{\frac{1}{0.951} - 0.34 \times \frac{0.5+5}{0.5+1}} = 0.30.$$

Given the estimated elasticity we can recover an approximate value for the average κ. From (2.31), we get an average value of $\hat{\kappa} = 0.93$. This value of κ implies that on average the top management team share 7% of the total firm sales.

Given our assumption of $\gamma > 0$, the value of $\frac{d\ln\pi}{d\ln y}$ found in the literature imposes a strong restriction on the possible values of σ. This can be seen in equation (2.30). Suppose we impose that $\frac{d\ln\pi}{d\ln y} < 1$ in this case we have $\gamma\kappa < \gamma\kappa\sigma$, since $\kappa > 0$ we then have $\sigma > 1$. This implies that a positive value of γ excludes the case of a Cobb-Douglas production function ($\sigma = 1$), an observation originally made in Rosen (1982). As mentioned earlier the formula for γ in equation (2.33) does not return a positive value of γ for all values of ε and σ. Given the estimated firm level

[21] Gabaix and Landier (2008) use EXECUCOMP data from 1992 to 2004 and select for each year the 1000 highest paid CEOs, using the total compensation variable TDC1 at year t which include salary, bonus, restricted stock granted and Black-Scholes value of stock-option granted. Then they regress the log of total compensation of the CEO in year t on the log of the firm's size proxies in year $t-1$, controlling for year and industry fixed effects. Similarly, using manufacturing firms in CompuStat (1994-2010), Alder (2012) finds the elasticity of CEOs with respect to employee size is 0.318 controlling for industry fixed effects.

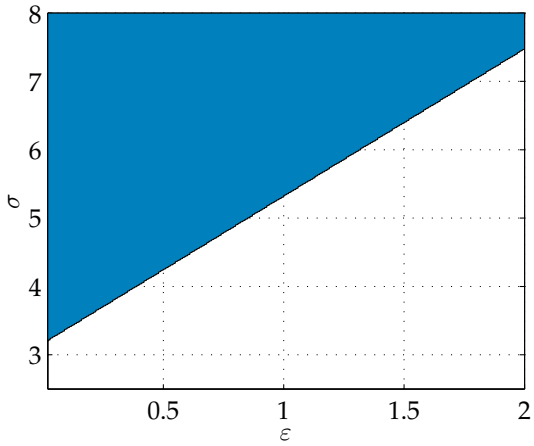


Figure 2.2: Admissible Region For (ε, σ).

elasticities, Figure 2.2 displays as a shaded region the permissible pairs of (ε, σ).

2.5.3 ESTIMATING a

The fact that managerial skills θ are unobservable makes the estimation of the the Pareto parameter a potentially challenging. This subsection shows how to recover the shape of the tail of the skill distribution using stylized facts on the distribution of firm sizes observed in data. Given Assumption 2.4.3 concerning the distribution of θ we have that given a realization $\{\theta^1, \ldots, \theta^N\}$ of managerial types the maximum likelihood estimate of a satisfies:[22]

$$\frac{1}{\hat{a}} = \frac{1}{N} \sum_{i=1}^{N} \left(\ln(\theta^i) - \ln(\underline{\theta}) \right), \tag{2.35}$$

where $\underline{\theta}$ is the minimum possible value of θ.

In our model the size distribution of firms reflects the skill distribution of managers. As a consequence, we can recover the estimate of a based on the observed distribution of $L(\theta)$. It is well documented that the distribution of firm size exhibits a Pareto distribution with tail index close to one.[23] If we let a_L denote the tail parameter of the Pareto distribution of firm size, the analogue to (2.35) yields

$$\frac{1}{\hat{a}_L} = \frac{1}{N} \sum_{i=1}^{N} \left(\ln L(\theta^i) - \ln(L(\underline{\theta})) \right). \tag{2.36}$$

From equation (2.26) in Lemma 2.5.1 we have that

$$\ln L(\theta) - \ln L(\underline{\theta}) = \left(1 + \frac{\gamma\sigma}{1-\hat{\kappa}} + \varepsilon \left(1 + \frac{\gamma}{1-\hat{\kappa}} \right) \right) (\ln(\theta) - \ln(\underline{\theta})),$$

[22] See, e.g., Malik (1970).
[23] See, among others, Simon and Bonini (1958), Ijiri and Simon (1964)

where we proxied $\kappa(\theta)$ with $\hat{\kappa}$. Substituting in (2.35) we get

$$\frac{1}{\hat{a}} = \frac{1}{1+\frac{\gamma\sigma}{1-\hat{\kappa}}+\varepsilon\left(1+\frac{\gamma}{1-\hat{\kappa}}\right)}\frac{1}{N}\sum_{i=1}^{N}\left(\ln L(\theta^i)-\ln L(\underline{\theta})\right). \tag{2.37}$$

Assuming that regularity conditions necessary for consistency of maximum likelihood estimates hold, (2.37) and (2.36) give the following result:

Proposition 2.5.2. *For any solution of the θ-manager's problem in (2.23) the following relationship holds:*

$$a = \underbrace{\left(1+\frac{\gamma\sigma}{1-\kappa}+\varepsilon\left(1+\frac{\gamma}{1-\kappa}\right)\right)}_{\frac{d\ln L}{d\ln\theta}}\times a_L, \tag{2.38}$$

From the above relationship we can recover a using information on a_L and our benchmark values for the rest of the parameters. Taking the estimate of the tail parameter of firm size from Axtell (2001), we set $a_L = 1.06$. Using this into equation (2.38) we get $a = 26.13$.

The above analysis linking the distribution of skills to the distribution of talent is an application of the lessons learned in Lucas (1978) and in Rosen (1982).[24] The implications of this identification step cannot be understated. Indeed, we see that the distribution of skills is significantly less skewed than the distribution of firm size (also refer to Section 2.7.2 looking at the distribution of income). Not recognizing this relationship between skills and observables could lead in over-estimating the thickness of the right tail of the skill distribution. In Saez (2001), for example, since wages are exogenously given, skills are identified via the distribution of income which leads to estimating a distribution of skills with tail parameter around 2. Looking back at our tax formula in equation (2.21) the inverse relationship between the optimal top tax rate and the tail parameter is clear. Hence, a low estimate of this parameter will mechanically lead to recommending high top tax rates. We come back to this point in Section 2.10.6.

To conclude this section, Table 2.2 summarizes the parameter calibration of the benchmark model.

We next compute the value for optimal taxes a the top.

2.6 OPTIMAL TOP INCOME TAX RATES

Substituting the parameters of the benchmark calibration from Table 2.2 into our top tax formula (2.21), we obtain that the optimal tax rate implied by our environment:

$$T'(\pi^*(\infty)) = \frac{1}{1+a\left[\frac{1}{\varepsilon}+1+\frac{\gamma}{1-\kappa}\left(\frac{1}{\varepsilon}+1-\rho\kappa\right)\right]^{-1}} = .324.$$

[24] Lucas (1978) focuses on the distribution of firm size, while Rosen (1982) focuses on the distribution of income.

Table 2.2: Benchmark Parameter Calibration.

Parameter	Symbol	Value
Frisch Elasticity	ε	0.5
Degree of substitutability of inputs	ρ	0.8
Elasticity of firm size w.r.t. sales	$\frac{d \ln L(\theta)}{d \ln y(\theta)}$	1.051
Elasticity of executive compensation w.r.t. sales	$\frac{d \ln \pi(\theta)}{d \ln y(\theta)}$	0.34
Scale-of-operations effect	γ	0.30
Share of labor costs to total sales	κ	0.93
Inverse of Pareto tail parameter of θ	a	26.13

Notes: ε and ρ are imposed exogenously. The elasticities are estimated from data following Section 2.5.2. γ, κ and a are respectively computed by (2.33), (2.31) and (2.38).

It is useful to benchmark the above result with two values. We begin with what we observe in data. We look at the March edition of the CPS from 2000 to 2010, and for every individual we compute federal and state taxes of labor income using the NBER TAXSIM calculator.[25] For the top 99$^{\text{th}}$ percentile we find an effective marginal federal tax rate of 33.5% and an effective marginal state tax rate of 5%. Saez et al. (2012) report a top 1% marginal rate of approximately 50% post 2000. With these bounds in mind our benchmark prescribes a lower tax rate than what we currently see in data. The prescription relies crucially on the estimated value of γ. To see this it is instructive to compare the result with the case in which $\gamma = 0$.[26] In this case $a_{\gamma=0} = 1.59$ and the corresponding tax rate is almost double our benchmark at 65.4%. In Section 2.7 we look at how top tax rate are affected by a and κ. We next proceed and study the effect of the elasticity of labor supply (ε) and the degree of substitutability across inputs (σ). In 2.3, as expected, we observe that tax rates are decreasing in ε (see also Figure 2.3(a)). When looking at changes of the elasticity of substitution between effort of the manager and of the worker we see that marginal taxes rates are increasing in σ. This is also more clearly displayed in Figure 2.3(b) where marginal taxes are displayed over a wider range of σ consistent with a positive value of γ as displayed in the shaded region of Figure 2.2. When changing values of either ε and σ given equation (2.33) and (2.38) we re-compute the values for γ and a. Table 2.3 displays the re-estimated values, also displayed are the values of marginal taxes when $\gamma = 0$. Compared to our benchmark marginal taxes are higher. However, the difference between the two (in relative terms) is decreasing as the manager labor supply becomes more inelastic.

[25] We drop individuals with negative income and labor income below \$100. Also dropped are individuals for which labor income is less than 60% of total income or more than 120% of total income. Tax rates are computed using the NBER TAXSIM calculator version 9.2. Rates reported are apply to the head of household inclusive of transfer received. Refer to Ales et al. (2014) for further details.

[26] In Section 2.10.6 we will show how the case with $\gamma = 0$ collapses our environment to the one studied in Saez (2001).

Table 2.3: Top Tax Rates.

	$\varepsilon = 0.25$		$\varepsilon = 0.50$		$\varepsilon = 0.75$	
σ	4.9	5.2	4.9	5.2	4.9	5.2
T' (%)	51.8%	51.9%	32.3%	32.6%	21.7%	22.1%
γ	0.16	0.13	0.35	0.24	2.09	0.63
a	13.5	11.7	29.3	21.8	176.5	57.3
$T'_{\gamma=0}$ (%)	79.1%	79.1%	65.4%	65.4%	55.7%	55.7%
$a_{\gamma=0}$	1.3	1.3	1.6	1.6	1.8	1.8

Notes: T' denotes optimal tax rate as imputed by (2.21). γ and a are respectively computed by (2.33) and (2.38). $T'_{\gamma=0}$ denote the optimal tax rates with the exogenous constraint of $\gamma = 0$.

2.7 ROBUSTNESS

In this section we revisit our optimal tax result by pursuing different quantitative strategies. In Section 2.7.1 we use information on the share of labor cost to total sales as opposed to the behavior of managerial compensation to sales. In Section 2.7.2 we determine the distribution of skills using the distribution of income rather than the distribution of firm sizes.

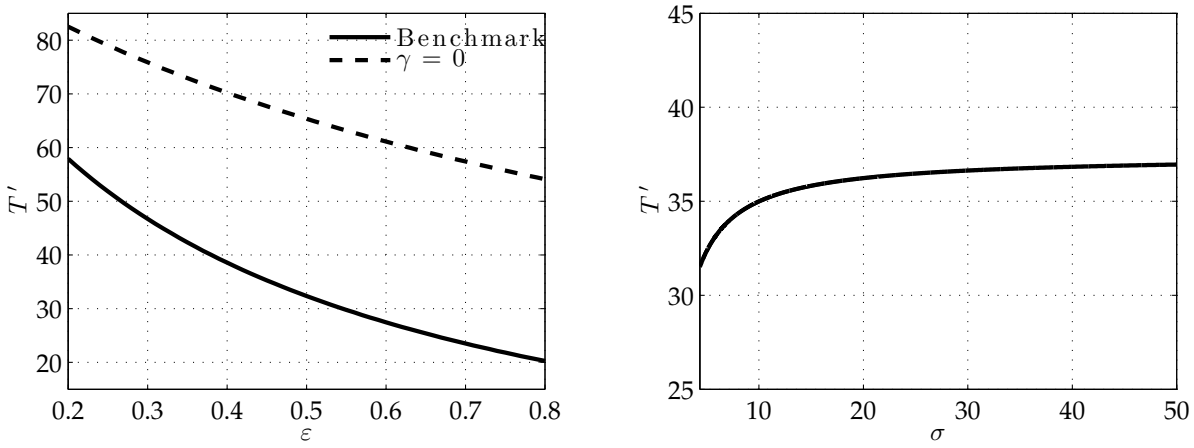
2.7.1 OPTIMAL TAXES USING κ FROM DATA

Our goal is to determine κ not via equation (2.31) as we have done in the previous section but instead estimating it directly from data. We use the Current Population Survey (CPS) administered by the US Census Bureau and the US Bureau of Labor Statistics. We look at years 1976 to 2012. We keep individuals between ages 25 to 65 and drop individuals not in the labor force or currently unemployed (for additional details on data and description of the sample refer to Ales et al. (2014)). We divide the population in two groups those whose occupation is describe as being: *Executive, Administrative, and Managerial Occupations* and those who are not (see Appendix 2.10.5 for additional details). For each group we compute total compensation. We then estimate κ as the total compensation of individuals not described as managers or individuals described as managers. Averaging over all years we get an average value of $\kappa = .812$. Next we back out γ using equation (2.32).

$$\gamma = \frac{\left(1 - \frac{d \ln L(\theta)}{d \ln y(\theta)}\right)(1-\kappa)(1+\varepsilon)}{\frac{d \ln L(\theta)}{d \ln y(\theta)}(1 - \kappa + \kappa\sigma + \varepsilon) - (\sigma + \varepsilon)} = 0.029. \tag{2.39}$$

The implied value for a is equal to 2.48. Finally going back to our tax formula in (2.21) we get $\bar{T}' = 57.5\%$.[27]

[27] Previous version of this draft used the NBER-CES Manufacturing Industry Database. With this data we estimated a value of $\kappa = 0.64$ with a confidence interval of $(0.580, 0.702)$. This lower value of κ is due to the fact that the dataset consider as managerial workers a large class of workers. From Berman et al. (1994) these workers include "those



(a) Varying ε.

(b) Varying σ.

Figure 2.3: Comparative Statics of Optimal Top Tax Rates.

2.7.2 ESTIMATING a FROM THE INCOME DISTRIBUTION

In Subsection 2.5.3 we estimated a using the distribution of firm sizes. In this section we proceed similarly but we focus instead on the distribution of incomes. From equation (2.28) in Lemma 2.5.1 we have that

$$\ln \pi(\theta) = \left(1 + \frac{\gamma}{1-\kappa(\theta)}\right)(1+\varepsilon)\ln\theta.$$

As before, approximating $k(\theta) = \hat{\kappa}$ and substituting in (2.35) we get

$$\frac{1}{a} = \frac{1}{\left(1+\frac{\gamma}{1-\hat{\kappa}}\right)(1+\varepsilon)}\frac{1}{N}\sum_{i=1}^{N}\ln\pi(\theta^i).$$

Assume that in the data income is distributed according to a Pareto distribution with tail parameter a_y. We then have $\frac{1}{a_y} = \frac{1}{N}\sum_{i=1}^{N}\ln\pi(\theta^i)$. Substituting in the above we have:

$$a = \left(1 + \frac{\gamma}{1-\hat{\kappa}}\right)(1+\varepsilon)a_y. \tag{2.40}$$

Taking $a_y = 2$ from Saez (2001) with the above estimates of κ and γ we get $a = 15.62$. In this case the top tax rate is equal to 44.4%.

engaged in supervision (above the working foreman level), installation and servicing of own product, sales, delivery, professional, technological, administrative etc."

2.8 OPTIMAL FIRM SIZE TAXATION

As emphasized in Section 2.4.1, typically the marginal product of labor is not equalized across firms at the optimum. This feature is necessary for incentive provision and, in our decentralization, translates into nonzero marginal taxes on firm size (see equation (2.15)). In this section we characterize optimal taxes on firm size in a calibrated example.

Unlike in previous sections where we focused on taxation at the top, here we compute firm size taxes over the entire skill distribution in order to study progressivity. Our calibration is as follows. The parameter controlling the right tail of the skill distribution is set to $a = 26.13$ as in our benchmark calibration. We assume that the skill distribution is Pareto-Lognormal[28] with $\theta \sim PlN(\zeta, \iota^2, a)$, and following Mankiw et al. (2009) we set $\zeta = 2.76$ and $\iota = 0.56$. The values for the parameters ε, ρ, and γ are taken from Table 2.2. To calibrate β (the share parameter on the production function), we use the definition of κ which implies:

$$\kappa(\theta) = \frac{MP_{L(\theta)}L(\theta)}{y(\theta)} = \frac{1}{1 + \frac{\beta}{1-\beta}\left(\frac{n\theta}{L}\right)^{\rho}}.$$

From NBER-CES Manufacturing Industry Database, the average team size defined by the number of production workers per non-production worker is estimated to be 3.58. Using this number to proxy $\left(\frac{n\theta}{L}\right)^{-1}$, the equation above and our benchmark value of $\kappa = 0.93$ together imply $\beta = 0.17$. Finally, the social welfare function is such that $\Psi(U) = U^{1/2}$.

Table 2.4 reports optimal marginal firm size taxes across firm size percentiles in our calibrated model. Two facts stand out. First, computed firm distortions are positive and economically significant. For the median firm, for example, these distortions raise the marginal labor cost by 1.75% above the wage. Second, the marginal tax on labor use is progressive in the range of firm sizes reported. In particular, T_L' increases by 30% between the 25^{th} percentile and the top, where it asymptotes at around 2%.

It is useful to benchmark the aforementioned prescriptions against actual firm size distortions in the data. As for the shape of such wedges, our model model generates progressive firm distortions, which is a property that is ubiquitous across countries.[29] When comparing levels, on the other hand, one faces a significant obstacle: in the real world, policies distorting firm sizes include a variety of labor regulations which cannot be easily summarized into a "tax equivalent" measure as in our model.[30] However, a recent study by Garicano et al. (2013) circumvents this problem by structurally estimating the tax equivalent of French labor regulations, which affect firms with more than 50 employees. Using a Lucas (1978) span of control model, the authors find that such labor legislation increase the cost of labor by around 1.3% of the wage, which is in

[28] That is, $\theta \stackrel{d}{=} \theta_1\theta_2$, where $\theta_1 \sim LN(\zeta, \iota^2)$ and $\theta_2 \sim P(a)$.

[29] See, e.g., Guner et al. (2008).

[30] Regulations on large firms range from penalties for not offering health insurance to their employees in the US, to rehiring unfairly dismissed workers in Italy. Other types of policies aim to promote smaller firms.

the same ballpark as the numbers in Table 2.4. From this perspective, our model can potentially rationalize both the shape and levels of existing firm size distortions in the data.

Table 2.4: Optimal Marginal Tax on Firm Size.

	Firm Size Percentile				
	25^{th}	50^{th}	75^{th}	90^{th}	99^{th}
$T'_L(wL)$	1.54%	1.75%	1.91%	2.01%	2.02%

2.9 CONCLUSION

The title of this paper is a reference to the thought provoking novel of Rand (1957).[31] In this dystopian novel what we would call "top income earners" reduce their labor effort in response to high taxes. In the novel, this action is described as having dramatic and long lasting effects for the economy. In this paper, we quantify such effects. Top income earners are modeled as managers whose effort and skill contribute–together with hired labor–to generate output. Our key finding is that tax rates should be substantially lower than what previous literature with unbounded distribution has found. On the other hand, current US marginal tax rate are close to what the normative benchmark prescribes. This result is robust to a large range of parameter specifications.

Methodologically, this paper highlights how to consider firm level data to determine key parameters relevant for income taxation. The logical next step involve taking a close look at the impact of managers on production. Two extensions come to mind: the first one is to determine the impact of a hierarchical organizations rather than a single manager technology for optimal taxes. The second extension is to consider the case in which managerial ability and labor productivity are drawn independently rather than being jointly determined by a unique skill parameter.

[31] The excellent survey of Slemrod (2000) also features a similar title.

2.10 APPENDIX

2.10.1 PROOF OF PROPOSITION 2.3.1

Define $M(\theta', \theta) \equiv c(\theta') - v(n(y(\theta'), L(\theta'), \theta))$. Incentive compatibility (2.4) requires that for all $\theta \in \Theta$, $M(\theta', \theta)$ attain a global maximum at $\theta' = \theta$. We start by characterizing local maxima of $M(\theta', \theta)$ at $\theta' = \theta$ using the following Lemma.

Lemma 2.10.1. *Let $M(\theta', \theta) \equiv c(\theta') - v(n(\theta', \theta))$ where $n(\theta', \theta) \equiv n(y(\theta'), L(\theta'), \theta)$. A local maximum of $M(\theta', \theta)$ at $\theta' = \theta$ is attained if and only if for all $\theta \in \Theta$:*

$$c'(\theta) - v'(n(\theta)) \left[n_y(\theta) y'(\theta) + n_L(\theta) L'(\theta) \right] = 0, \tag{2.41}$$

and

$$\begin{aligned} &y'(\theta) \left[v''(n(\theta)) n_y(\theta) n_\theta(\theta) + v'(n(\theta)) n_{\theta y}(\theta) \right] + \\ &+ L'(\theta) \left[v''(n(\theta)) n_L(\theta) n_\theta(\theta) + v'(n(\theta)) n_{\theta L}(\theta) \right] \leq 0. \end{aligned} \tag{2.42}$$

where $n(\theta, \theta) = n(\theta)$ and $n_i(\theta, \theta) = n_i(\theta)$ for $i = y, L, \theta, y\theta, L\theta$.

Proof. The first order condition for $\theta' = \theta$ to be a local maximum of $M(\theta', \theta)$ is $M_1(\theta, \theta) = 0$.[32] This is equivalent to (2.41). Differentiating the first order condition $M_1(\theta, \theta) = 0$ with respect to θ gives $M_{11}(\theta) + M_{12}(\theta) = 0$. Hence, the second order condition $M_{11}(\theta) \leq 0$ can be written as $-M_{12}(\theta) \leq 0$, which gives (2.42). □

We now go back to the proof of Proposition 2.3.1. The proof follows standard arguments.

Proof. We want to show that if (2.41) and (2.42) hold, then $M_1(\theta', \theta)$ has the sign of $(\theta - \theta')$. This implies that $M(\theta', \theta)$ attains a global maximum at $\theta' = \theta$ so that (2.4) holds. First note that

$$M_1(\theta', \theta) = c'(\theta') - v'(n(\theta', \theta)) \left[n_y(\theta', \theta) y'(\theta') + n_L(\theta', \theta) L'(\theta') \right]. \tag{2.43}$$

We also have that (2.41) evaluated at θ' gives

$$c'(\theta') = v'(n(\theta')) \left[n_y(\theta') y'(\theta') + n_L(\theta') L'(\theta') \right]. \tag{2.44}$$

Using (2.44) into (2.43) gives

$$M_1(\theta', \theta) = J(\theta', \theta') - J(\theta', \theta), \tag{2.45}$$

where $J(\theta', \theta) \equiv v'(n(\theta', \theta)) \left[n_y(\theta', \theta) y'(\theta') + n_L(\theta', \theta) L'(\theta') \right]$. Differentiating with respect to the second argument:

$$J_2(\theta', \theta') = y'(\theta') \left[v''(n(\theta')) n_y(\theta') n_\theta(\theta') + v'(n(\theta')) n_{\theta y}(\theta') \right] +$$

[32] The subscript $i = 1, 2$ denote derivative with respect to respectively the first or second argument

$$L'(\theta')\left[v''(n(\theta'))n_L(\theta')n_\theta(\theta')+v'(n(\theta'))n_{\theta L}(\theta')\right];$$

From (2.42) we have that $J_2(\theta',\theta')\leq 0$. Then (2.45) implies that $M_1(\theta',\theta)\geq 0$ if and only if $\theta'\leq\theta$. Finally (2.8) is obtained by combining (2.41) and (2.42). This completes the proof. □

2.10.2 PROOFS OF SECTION 2.4

We compute Pareto optimal allocation by solving the optimal control problem (PO-FOC) where $y(\theta)$ and $L(\theta)$ are the controls and $U(\theta)$ is the state variable. After integrating by parts, the Lagrangian to the planner's problem is (suppressing dependencies with respect to θ, y and L):

$$\mathcal{L}=\Psi(c^w)+\int\Psi(U)dF-\int[\mu' U-\mu v'(n)n_\theta]d\theta+\lambda^r\int[y-c^w-U-v(n)]dF-\lambda^l\int[L-1]dF,$$

where λ^r is the multiplier on (2.6), λ^l is the multiplier on (2.7) and $\mu(\theta)$ is the costate on (2.5) that also satisfies the boundary conditions $\mu(\underline{\theta})=\lim_{\theta\to\bar{\theta}}\mu(\theta)=0$. It is straightforward to show that all of these multipliers are positive.

Optimality conditions with respect to the controls y, L are, respectively,

$$\begin{aligned}\lambda^r(1-v'(n)n_y)f+\mu(v''(n)n_yn_\theta+v'n_{\theta y}) &= 0, &(2.46)\\ -\lambda^l f-\lambda^r v'(n)n_L f+\mu(v''(n)n_Ln_\theta+v'n_{\theta L}) &= 0 &(2.47)\end{aligned}$$

and the costate equation is

$$\mu'=(\Psi'(U)-\lambda^r)f. \tag{2.48}$$

PROOF OF PROPOSITION 2.4.1

Let (y,L,θ) be such that $n(y,L,\theta)=\bar{n}$, where $\bar{n}$ is a given level of effort. By applying the implicit function theorem we have that

$$y_L=-n_L/n_y. \tag{2.49}$$

Rearranging the first order conditions (2.46) and (2.47) we get

$$n_y[\lambda^r v'f-\mu v''n_\theta]=\lambda^r f+\mu v'n_{\theta y},$$

and

$$n_L[\lambda^r v'f-\mu v''n_\theta]=-\lambda^l f+\mu v'n_{\theta L},$$

so that

$$y_L=-\frac{n_L}{n_y}=\frac{\lambda^l f-\mu v'n_{\theta L}}{\lambda^r f+\mu v'n_{\theta y}}. \tag{2.50}$$

Using (2.15), (2.50) and substituting the expression for $w=\lambda^l/\lambda^r$ we derive equation (2.17).

To prove Part 2, suppose that $n_\theta = f(n)$ for all θ. It follows that for all θ

$$\frac{n_{\theta L}}{n_{\theta y}} = \frac{n_L}{n_y}. \tag{2.51}$$

Rearranging (2.50) and substituting (2.51) we get

$$-\frac{n_{\theta L}}{n_{\theta y}}\lambda^r = \lambda^l,$$

so simplifying and substituting (2.51) we get $-n_L/n_y = \lambda^l/\lambda^r$, so that $T'_L = 0$.

PROOF OF PROPOSITION 2.4.2

Integrating (2.48) between θ and $\overline{\theta}$ and using the transversality condition we get

$$\mu(\theta) = \int_\theta^{\overline{\theta}} \left(\lambda^r - \Psi'(U(\theta))\right) f(\theta) d\theta. \tag{2.52}$$

Evaluating (2.52) at $\underline{\theta}$ gives

$$\lambda^r = \int_\Theta \Psi'(U(\theta)) f(\theta) d\theta. \tag{2.53}$$

Let $D(\theta) \equiv \frac{1}{1-F(\theta)} \int_\theta^\infty \Psi'\left(U(\theta)\right) dF(\theta)$ so that $D(\underline{\theta}) = \lambda^r$. Substituting into (2.52) we obtain

$$\mu(\theta) = (1 - F(\theta)) \left(D(\underline{\theta}) - D(\theta)\right), \tag{2.54}$$

from (2.46)

$$-\mu n_y v' \left[\frac{v''}{v'} n_\theta + \frac{n_{\theta y}}{n_y}\right] = \lambda^r [1 - v'(n) n_y] f,$$

substitute (2.54)

$$(1 - F(\theta)) \left(D(\underline{\theta}) - D(\theta)\right) n_y v' \left[-\frac{v''}{v'} n_\theta - \frac{n_{\theta y}}{n_y}\right] = \lambda^r [1 - v'(n) n_y] f.$$

By Assumption 2.4.1 and (2.53) we get

$$\frac{(1 - F(\theta))}{\theta f(\theta)} \left(1 - \frac{D(\theta)}{D(\underline{\theta})}\right) \left[-\frac{1}{\varepsilon}\frac{n_\theta}{n}\theta - \frac{n_{\theta y}}{n_y}\theta\right] = \frac{[1 - v'(n) n_y]}{v'(n) n_y},$$

so using (2.18) yields

$$\frac{T'}{1 - T'} = \frac{1 - F(\theta)}{\theta f(\theta)} \left(1 - \frac{D(\theta)}{D(\underline{\theta})}\right) \left[-\frac{1}{\varepsilon}\frac{n_\theta}{n}\theta - \frac{n_{\theta y}}{n_y}\theta\right]. \tag{2.55}$$

Next we write the partial derivatives of n in (2.55) in terms of partial derivatives of y. Let (n, L, θ) be such that $y(n, L, \theta) = \bar{y}$, where $\bar{y}$ is a given level of output. The implicit function theorem gives

$$n_L = -\frac{y_L}{y_n}, \tag{2.56}$$

and

$$n_\theta = -\frac{y_\theta}{y_n}. \tag{2.57}$$

Combining (2.49) and (2.56) gives $n_y(y, L, \theta) = 1/y_n(n(y, L, \theta), L, \theta)$. By differentiating both sides with respect to θ we have $n_{\theta y} = -\frac{y_{nn} n_\theta + y_{n\theta}}{y_n^2}$, which implies

$$-\frac{n_{\theta y}}{n_y}\theta = \left(\frac{y_{nn}}{y_n} n\right)\left(\frac{n_\theta}{n}\theta\right) + \frac{y_{n\theta}}{y_n}\theta. \tag{2.58}$$

Substituting (2.57) and (2.58) into (2.55) gives the result.

PROOF OF COROLLARY 2.4.1

Under this technology, it is straightforward to verify that

$$\frac{y_\theta}{y_n} \cdot \frac{\theta}{n} = 1, \qquad \frac{y_{nn}}{y_n} n = \frac{H_{11}}{H_1}\theta n, \qquad \frac{y_{n\theta}}{y_n}\theta = \frac{H_{11}}{H_1}\theta n + 1.$$

Substituting the above in (2.19) yields (MDS).

PROOF OF COROLLARY 2.4.2

Denote with $g(\theta) = \theta^\gamma$. Given Assumption 2.4.2 we have

$$n(y, L, \theta) = \left[\frac{1}{\beta}\left(\frac{y}{\theta g(\theta)}\right)^\rho - \frac{1-\beta}{\beta}\left(\frac{L}{\theta}\right)^\rho\right]^{\frac{1}{\rho}}. \tag{2.59}$$

Taking derivatives from (2.59) we obtain

$$n_\theta(y, L, \theta) = \left(-\frac{1}{\theta}\right)\left[\frac{1}{\beta}\left(\frac{y}{\theta g(\theta)}\right)^\rho - \frac{1-\beta}{\beta}\left(\frac{L}{\theta}\right)^\rho\right]^{\frac{1}{\rho}-1}\left[\frac{1}{\beta}\left(\frac{y}{\theta g(\theta)}\right)^\rho (1+\gamma) - \frac{1-\beta}{\beta}\left(\frac{L}{\theta}\right)^\rho\right],$$

using the above and (2.59) we get

$$\frac{n_\theta(y, L, \theta)}{n(y, L, \theta)} = -\frac{1}{\theta}\frac{\left(\frac{y}{g(\theta)}\right)^\rho (1+\gamma) - (1-\beta)L^\rho}{\left(\frac{y}{g(\theta)}\right)^\rho - (1-\beta)L^\rho}. \tag{2.60}$$

Also

$$n_y(y, L, \theta) = \left[\frac{1}{\beta}\left(\frac{y}{\theta g(\theta)}\right)^\rho - \frac{1-\beta}{\beta}\left(\frac{L}{\theta}\right)^\rho\right]^{\frac{1}{\rho}-1}\frac{1}{\beta y}\left(\frac{y}{\theta g(\theta)}\right)^\rho, \tag{2.61}$$

$$n_{\theta y}(y,L,\theta) = \left(-\frac{1-\rho}{\theta}\right)\frac{1}{\beta y}\left(\frac{y}{\theta g(\theta)}\right)^{\rho}\left[\frac{1}{\beta}\left(\frac{y}{\theta g(\theta)}\right)^{\rho} - \frac{1-\beta}{\beta}\left(\frac{L}{\theta}\right)^{\rho}\right]^{\frac{1}{\rho}-2} \times$$
$$\left[\frac{1}{\beta}\left(\frac{y}{\theta g(\theta)}\right)^{\rho}(1+\gamma) - \frac{1-\beta}{\beta}\left(\frac{L}{\theta}\right)^{\rho}\right] - \frac{\rho}{\theta}\left[\frac{1}{\beta}\left(\frac{y}{\theta g(\theta)}\right)^{\rho} - \frac{1-\beta}{\beta}\left(\frac{L}{\theta}\right)^{\rho}\right]^{\frac{1}{\rho}-1}\frac{1}{\beta y}\left(\frac{y}{\theta g(\theta)}\right)^{\rho}(1+\gamma).$$

The two above imply:

$$\frac{n_{\theta y}(y,L,\theta)}{n_y(y,L,\theta)} = -\frac{1}{\theta}\left[(1-\rho)\frac{\left(\frac{y}{g(\theta)}\right)^{\rho}(1+\gamma) - (1-\beta)L^{\rho}}{\left(\frac{y}{g(\theta)}\right)^{\rho} - (1-\beta)L^{\rho}} + \rho(1+\gamma)\right]. \tag{2.62}$$

By Assumption 2.4.2 we have

$$\left(\frac{y(\theta)}{g(\theta)}\right)^{\rho} = \beta(\theta n)^{\rho} + (1-\beta)L^{\rho}.$$

Then

$$\frac{\left(\frac{y}{g(\theta)}\right)^{\rho}(1+\gamma) - (1-\beta)L^{\rho}}{\left(\frac{y}{g(\theta)}\right)^{\rho} - (1-\beta)L^{\rho}} = 1 + \gamma\left(1 + \frac{1-\beta}{\beta}\left(\frac{L}{\theta n}\right)^{\rho}\right) \tag{2.63}$$

Also, Assumption 2.4.2 implies:

$$\frac{\kappa(\theta)}{1-\kappa(\theta)} = \frac{1-\beta}{\beta}\left(\frac{L}{\theta n}\right)^{\rho}, \tag{2.64}$$

where $\kappa(\theta) \equiv y_L(\theta)L(\theta)/y(\theta)$ denotes the *share of labor costs to total sales* for manager θ. Using (2.64) in (2.63) gives

$$\frac{\left(\frac{y}{g(\theta)}\right)^{\rho}(1+\gamma) - (1-\beta)L^{\rho}}{\left(\frac{y}{g(\theta)}\right)^{\rho} - (1-\beta)L^{\rho}} = 1 + \gamma\left(1 + \frac{\kappa(\theta)}{1-\kappa(\theta)}\right). \tag{2.65}$$

Using (2.60), (2.62) and (2.65) into (2.55) gives the result.

2.10.3 PROOFS OF SECTION 2.5

PROOF OF LEMMA 2.5.1

We start by establishing a the following result (henceforth we suppress the arguments of h, h', h''):

Claim 1.

$$\frac{1}{\sigma} = -\left(\frac{\theta n}{L}\right)\frac{h''}{h'}\frac{1}{\kappa}. \tag{2.66}$$

Proof. From the definition of κ we can write

$$1-\kappa=\frac{\theta n}{L}\frac{h'}{h}. \tag{2.67}$$

Let $f(\theta n, L)=Lh(\theta n/L)$. Define the elasticity of substitution between θn and L as

$$\sigma\equiv-\frac{d\ln(L/\theta n)}{d\ln(f_2/f_1)}.$$

By definition of f we have $f_1=h'$ and $f_2=h-\frac{\theta n}{L}h'$ which implies $\frac{f_2}{f_1}=\frac{h}{h'}-\frac{\theta n}{L}$. Therefore

$$\frac{1}{\sigma}=-\frac{d\ln(f_2/f_1)}{d\ln(L/\theta n)}=-\left(\frac{h}{h'}-\frac{\theta n}{L}\right)^{-1}\frac{L}{\theta n}\frac{d\left(\frac{h}{h'}-\frac{\theta n}{L}\right)}{d(L/\theta n)}.$$

Differentiating and re-arranging, we get:

$$\frac{1}{\sigma}=-\left(\frac{\theta n}{L}\right)\frac{h''}{h'}\left(1-\frac{\theta n}{L}\frac{h'}{h}\right)^{-1}.$$

Substituting (2.67) in the above we obtain the result. □

We now move to the proof of Lemma 2.5.1. As notation let $g=\theta^{\gamma}$ and $g'=\gamma\theta^{\gamma-1}$.

Proof. Differentiating (2.24) and (2.25) we get

$$\frac{d\ln L}{d\ln\theta}-\frac{d\ln n}{d\ln\theta}=1-\theta\frac{g'}{g}\left(\frac{L}{\theta n}\right)^2\frac{1}{h''}\left(h-\frac{\theta n}{L}h'\right), \tag{2.68}$$

$$\frac{n\theta}{L}gh''\frac{d\ln L}{d\ln\theta}+\left[\frac{n}{\theta}\frac{v''}{(1-\tau)\chi}-\frac{\theta n}{L}gh''\right]\frac{d\ln n}{d\ln\theta}=g'\theta h'+gh'+gh''\frac{n\theta}{L}. \tag{2.69}$$

Combining (2.68) and (2.69),

$$\frac{n}{\theta}\frac{v''}{(1-\tau)\chi}\frac{d\ln n}{d\ln\theta}+\frac{n\theta}{L}gh''\left(1+\frac{\gamma\sigma}{1-\kappa}\right)=g'\theta h'+gh'+gh''\frac{n\theta}{L},$$

where we applied (2.66), (2.67) and the definition of γ on the last term. Further rearranging (2.66) and (2.67) gives

$$\frac{v''}{(1-\tau)\chi}\frac{n}{\theta}\frac{d\ln n}{d\ln\theta}=gh'\left(1+\frac{\gamma}{1-\kappa}\right). \tag{2.70}$$

From the first order condition (2.25) we have

$$\frac{d\ln n}{d\ln\theta}=\varepsilon\left(1+\frac{\gamma}{1-\kappa}\right), \tag{2.71}$$

where we used that $\frac{v'}{v''n} = \varepsilon$. Plugging (2.71) into (2.68),

$$\frac{d\ln L}{d\ln\theta} = \varepsilon\left(1+\frac{\gamma}{1-\kappa}\right)+1-\gamma\left(\frac{L}{\theta n}\right)^2\frac{1}{h''}\left(h-\frac{\theta n}{L}h'\right).$$

Then applying (2.66) and (2.67) and rearranging gives (2.26).

Now we obtain equation (2.27). From (2.22) we have

$$\begin{aligned}\frac{d\ln y}{d\ln\theta} &= \frac{d\ln g(\theta)}{d\ln\theta}+\frac{d\ln L}{d\ln\theta}+\frac{d\ln h\left(\frac{\theta n}{L}\right)}{d\ln\theta}\\ &= \gamma+\frac{d\ln L}{d\ln\theta}+\frac{\theta n}{L}\frac{h'}{h}\left(1+\frac{d\ln n}{d\ln\theta}-\frac{d\ln L}{d\ln\theta}\right).\end{aligned}$$

Substituting (2.67) in the above gives

$$\frac{d\ln y}{d\ln\theta} = \gamma+\kappa\frac{d\ln L}{d\ln\theta}+(1-\kappa)\left(1+\frac{d\ln n}{d\ln\theta}\right).$$

So using (2.26) and (2.71) into the above expression gives (2.27). Finally, we derive equation (2.28). Profits are given by

$$\pi(\theta) = y(\theta)-wL(\theta).$$

Then

$$\frac{d\ln\pi(\theta)}{d\ln\theta} = \frac{d\ln y(\theta)}{d\ln\theta}\frac{y(\theta)}{\pi(\theta)}-w\frac{L(\theta)}{\pi(\theta)}\frac{L(\theta)}{\pi(\theta)},$$

or

$$\frac{d\ln\pi(\theta)}{d\ln\theta} = \frac{d\ln y(\theta)}{d\ln\theta}\frac{1}{1-\kappa(\theta)}-w\frac{L(\theta)}{\pi(\theta)}\frac{\kappa(\theta)}{1-\kappa(\theta)}, \tag{2.72}$$

where $\kappa = wL/y$. Substituting (2.26) and (2.27) into (2.72) and rearranging gives (2.28).

□

2.10.4 FIRM DISTORTIONS AND TAX ELASTICITIES

In this section we show that if the are no firm level distortions then a wage rate exist for the effective effort of the manager and the income elasticity of the after tax rate equals the Frish elasticity of labor supply.

We first show that at the optimum is possible to write the income of managers of type θ as $\pi(\theta) = \omega(\theta)n$ where $\omega(\theta, w)$ is the wage of managers of type θ exercising effort n. The first order condition with respect to L is: $\theta^\gamma H_L(\theta n, L) = w$. Since H_L is a homogeneous of degree zero function we have $\theta^\gamma H_L(\theta n/L, 1) = w$ so that $\frac{\theta n}{L} = H_L^{-1}\left(\frac{w}{\theta^\gamma}, 1\right)$. This relationship implies that for a given θ and w the relationship between θn and L is linear. Define $m(\theta, w) = 1/H_L^{-1}\left(\frac{w}{\theta^\gamma}, 1\right)$. So that $L = m(\theta, w)\theta n$. Substituting in the expression for profits we have: $\pi(\theta, n) = \theta^\gamma H(\theta n, m(\theta, w)\theta n) -$

$m(\theta, w)w\theta n$. Since H is homogeneous of degree one we have:

$$\pi(\theta, n) = \left[\theta^{\gamma+1} H(1, m(\theta, w)) - wm(\theta, w)\theta\right] n = \omega(\theta, w)n.$$

We can now write the problem of the manager as:

$$\max c(\theta) - v(n(\theta)) \quad s.t. \quad c(\theta) = (1-\tau)\omega(\theta, w)n.$$

First order conditions of the above problem can be written as $n(\theta) = (v')^{-1}[(1-\tau)\omega(\theta, w)]$, so that:

$$\frac{\partial n}{\partial(1-\tau)} = \frac{1}{v''(n(\theta))} \cdot \omega(\theta, w) = \frac{v'(n(\theta))}{v''(n(\theta))} \cdot \frac{1}{(1-\tau)}, \tag{2.73}$$

where the second inequality follows from the first order condition. Substituting (2.73):

$$e \equiv \frac{\partial \log \omega(\theta, w)n}{\partial \log(1-\tau)} = \frac{\partial n(\theta)}{\partial(1-\tau)} \cdot \frac{1-\tau}{n(\theta)} = \varepsilon$$

This analysis would not apply in the case of a firm being subject to distortionary taxes or if the size of the firm where to be fixed L.

2.10.5 OCCUPATIONS USED FROM CPS

In CPS we identify as a manager individuals for which their 1990 occupation code is within the "*Executive, Administrative, and Managerial Occupations*". In detail:[33]

Table 2.5: Occupational Codes Used For Managerial Designation

Occupation code	Description
003	Legislators
004	Chief executives and public administrators
007	Financial managers
008	Human resources and labor relations managers
013	Managers and specialists in marketing, advertising, and public relations
014	Managers in education and related fields
015	Managers of medicine and health occupations
016	Postmasters and mail superintendents
017	Managers of food-serving and lodging establishments
018	Managers of properties and real estate
019	Funeral directors
021	Managers of service organizations
022	Managers and administrators

[33]Data is taken from: Miriam King, Steven Ruggles, J. Trent Alexander, Sarah Flood, Katie Genadek, Matthew B. Schroeder, Brandon Trampe, and Rebecca Vick. Integrated Public Use Microdata Series, Current Population Survey: Version 3.0. [Machine-readable database]. Minneapolis: University of Minnesota, 2010.

2.10.6 RELATIONSHIP WITH SAEZ (2001)

In Section 2.6 we compare our benchmark taxation result with the case in which $\gamma = 0$ thus removing any span of control feature from the environment. We will next show that setting $\gamma = 0$ collapses our environment to the one in Saez (2001). To see this consider equation (2.21) and, using (2.40), substitute a the tail parameter for the distribution of types with a_y the tail parameter for the distribution of income. We get:

$$T' = \frac{1}{1 + a_y \left(1 + \frac{\gamma}{1-\kappa}\right) \varepsilon \left[1 + \frac{\gamma}{1-\kappa}\left(1 - \frac{\rho\kappa\varepsilon}{1+\varepsilon}\right)\right]^{-1}}. \tag{2.74}$$

It is immediate that in the case with no scale-of-operations effect ($\gamma = 0$) or in the the Cobb-Douglas case ($\rho = 0$) then we get $T' = 1/(1 + a_y\varepsilon)$ as in Saez (2001) and Diamond and Saez (2011).

Equation (2.74) allows us to understand further the forces present in this environment compared to the benchmark case of Saez. As discussed in Diamond and Saez (2011) marginal taxes at the top can be understood looking at the distribution of types (a) and the income elasticity of the after tax rate $e \equiv \frac{\partial \log \pi(\theta)}{\partial \log(1-\tau)}$. In equation (2.40) we see how the presence of the scale-of-operations effect creates a wedge between the distribution of income and the distribution of managerial type. This is a force for lower taxes (since it points towards a higher value of a). At the same time firm level distortions that emerge with the scale-of-operations effect (as long as $\rho \neq 0$) generate a lower response of income to taxes.[34] This is a force for higher taxes. In the two cases discussed above, these two forces cancel each other perfectly. This is, however, not true in general as our computed example has demonstrated.

Remark 2.10.1. *The benchmark environment we have taken to the data will feature a lower income elasticity of the after tax rate than the one documented in the data (see, for example Saez et al. (2012) and Piketty et al. (2014)). However this elasticity, in our environment, is endogenous to policy. Firm level distortions having a particular strong effect. Indeed (as we show in Appendix 2.10.4) absent firm size distortions, as we think is the case in US data, the value of this elasticity will be given by the Frisch elasticity of labor supply. Hence in this case the value of "e"generated by the model would be consistent with the one estimated in data.*

[34] In Appendix 2.10.4 we show that absent firm level distortions the income elasticity of the after tax rate equals the Frish elasticity of labor supply.

Chapter 3

Regional State Capacity and the Optimal Degree of Fiscal Decentralization

joint with Martín Besfamille

3.1 INTRODUCTION

In many developed and developing countries, it is common that tax and expenditure assignments to subnational or regional governments be unbalanced. In particular, regional governments are often in charge of delivering local public services, but cannot raise the required revenues to finance these expenditures.[1] These so called "vertical fiscal imbalances" should either be covered by centrally provided transfers,[2] or just bypassed by decentralizing tax powers to regional governments.

The first alternative is widely used in practice and gives rise to an institutional setting defined as *partial decentralization*.[3] As shown by Wildasin (1997) and Goodspeed (2002), under this scenario regional governments may face "soft budget constraints,"[4] which can create negative externalities across regions and induce excessive spending or borrowing.[5] To cope with this problem, a vast literature in public finance has put forward different institutional mechanisms so that regional governments face "hard budget constraints."[6] One mechanism that has attracted significant

[1] Eyraud and Lusinyan (2013) report that across OECD countries, the average share of subnational government expenditure not financed through own revenues was 40 percent between 1995 and 2005. In Belgium and Mexico these shares climb to 60 and 83 percent, respectively. Corbacho et al. (2013) document that vertical fiscal imbalances in Latin America are the highest among developing nations.

[2] See Boadway and Shah (2007).

[3] This term has been coined by Brueckner (2008).

[4] According to Kornai et al. (2004), "*A budget-constrained organization faces a hard budget constraint as long as it does not receive support from other organizations to cover its deficit and is obliged to reduce or cease its activity if the deficit persists. The soft budget constraint phenomenon occurs if one or more supporting organizations are ready to cover all or part of the deficit*" (page 1097).

[5] Among others, Pettersson-Lidbom (2010) confirms this theoretical result by estimating that between 1979 and 1992, Swedish local governments increased their debt by more than 20 percent when they expected to receive future bailouts.

[6] See, among others, Rodden et al. (2003), Oates (2005) and Weingast (2009).

attention is the complete decentralization of tax powers to subnational governments. In a seminal contribution, Qian and Roland (1998) argue that such a system of *full decentralization* gives rise to tax competition, which in turn raises the perceived marginal costs of public funds at the regional level. But while this regime boosts fiscal discipline, hard budget constraints may also lead to underprovison of local projects, as argued by Besfamille and Lockwood (2008).

The goal of this paper is to provide a novel framework for comparing partial and full decentralization from a normative perspective. Our central departure relative to previous analyzes is to model local governments as being indexed by their level of *state capacity*, defined by Besley and Persson (2010) as the "state's ability to implement a range of policies." This institutional ingredient is of central importance for the study of the optimal degree of fiscal decentralization. In fact, a number of descriptive reviews of decentralization reforms (e.g., Bird (1995), Litvack et al. (1998)) and recent quantitative evaluations of such processes (e.g., Loayza et al. (2014)) argue that the benefits accruing from a decentralized form of government crucially hinge on these abilities. Or, more generally, Prud'homme (1995) and Bardhan and Mookherjee (2006), state that models formalizing pro-decentralization arguments usually ignore key institutional features of local governments.

We consider an environment in which regional governments decide whether or not to provide a discrete, local public good or project. The initial cost of the project is covered by the regional government's financial resources. If the project is initiated, it is carried out by regional bureaucracies, and there is a probability that the project is finished in due time and generates a social benefit greater than the initial cost. With the complementary probability, the project is delayed and needs a second round of financing to be completed. In this case, it generates a social benefit to the region, but lower than when the project is carried out on time. In the partially decentralized regime, no regional government has tax powers to refinance its incomplete project but the central government can bail out regions. Bailouts are financed via a uniform national tax on local capital, which is imperfectly mobile. Under full decentralization, on the other hand, regional governments have to refinance incomplete projects through a tax on capital invested in their jurisdiction, in a context of tax competition. Equilibrium outcomes under partial and full decentralization can be inefficient, as these regimes can generate overprovision or underprovision of projects, respectively.

Crucially, we make a distinction between two dimensions of local state capacity; namely, *administrative* and *fiscal*. The former measures the ability of subnational governments to produce and deliver public goods and services, and it is proxied by the probability that a project is finished on time. The latter gauges the capacity to raise revenues through local taxes, which is modeled as the fraction of the potential tax base that end up as fiscal revenues of the subnational governments. The model is symmetric: all regional governments have the same level of state capacity, and all costs and benefits are identical across regions. However, outcomes can be different ex post depending on the length of the projects.

The main results are the following. First, when the level of regional fiscal capacity that prevails in the federation is sufficiently low, partial decentralization dominates. Intuitively, refinancing

incomplete projects under full decentralization is too costly, regardless of the level of regional administrative capacity. Second, and more interestingly, we show that full decentralization dominates when combining high levels of fiscal capacity with low levels of administrative capacity. Essentially, although many projects remain incomplete, distortionary refinancing under full decentralization is unlikely. Thus, expected distortions under full decentralization are lower than under partial decentralization. This finding contradicts the views of certain policy proposals suggesting that high levels of administrative capacity are necessary for successful decentralization reforms (see, e.g., Bird (1995)), and it is illustrated using international data in the body of the paper.

We then undertake a comparative statics analysis. We show that when the regional capital stock increases, full decentralization dominates more frequently. Conversely, partial decentralization dominates more often when the highest possible benefit of the project rises. In addition, we show that when the number of regions increases, partial decentralization dominates in a parameter area where full decentralization was initially the optimal regime, and vice versa. Finally, we evaluate the robustness of our results by developing an extension that incorporates distortionary national taxation.

The layout of the remainder of the paper is as follows. Section 3.2 presents the model, and Section 3.3 describes the efficient benchmark. In Section 3.4 we analyze partial decentralization outcomes. Section 3.5 studies the equilibrium under full decentralization. Section 3.6 contains the normative comparison between partial and full decentralization, comparative statics and robustness results. Section 3.7 discusses the related literature. Section 3.8 concludes. The main proofs are contained in the Appendix. Additional proofs and derivations are relegated to the Online Appendix.

3.2 THE MODEL

3.2.1 PRELIMINARIES

The economy lasts for three periods $t = \{1, 2, 3\}$ and is composed of $L \geq 2$ regions. Each region $\ell \in \{1, ..., L\}$ has a continuum of measure 1 of risk-neutral, immobile residents, each of whom has an endowment κ of capital. In the last period, each resident derives utility from consumption of a numéraire good, produced in every region by competitive firms that operate a constant returns technology. Capital is the only input and units are chosen so that one unit of capital produces one unit of output. Following Persson and Tabellini (1992), we assume that capital is mobile between regions, but at a cost. Specifically, a resident of a region that invests f units of capital in other region(s) incurs a mobility cost $f^2/2$. As we explain below, residents may also benefit from a discrete local public good, or project.

There are two levels of government: central and regional. Throughout the paper, we assume that both levels of government are benevolent and choose policies so as to maximize the sum of utilities of their residents. For simplicity, there is no discounting of future payoffs.

3.2.2 TIMING

The order of events is as follows. At $t = 1$, a political body (e.g., a Congress) chooses between partial (PD) and full decentralization (FD). These institutional regimes rule all fiscal interactions between the central and regional governments, in a way specified below.

At the beginning of $t = 2$, Nature draws the cost c of the projects according to a strictly positive probability density function $h(c)$ on $[0, b]$. Based on this cost, regional governments choose whether or not to initiate a project in their region. This decision is denoted by $i_\ell \in \{I, NI\}$, where I (NI) stands for initiation (not initiation). Regional governments have just enough resources to fund the initial investment c. If initiated, a project is carried out by the regional bureaucracy. With an exogenous probability $\pi \in [0, 1]$, a project generates a social benefit $B > 0$ for all residents of the region at the end of the current period.[7] With probability $(1 - \pi)$, the project remains incomplete and yields no benefit during this period.[8] Projects' realizations are the result of independent draws from the probability distribution $(\pi, 1 - \pi)$, and are observable.

At $t = 3$, central or regional governments, depending on the institutional regime in place, decide whether to shut down or continue incomplete projects. In the last case, a project requires an additional input of c of the consumption good to be completed.

Under partial decentralization, the central government decides on refinancing incomplete projects, through a uniform tax τ on capital, collected by the national tax authority.[9] Under full decentralization, each regional government decides whether to refinance its incomplete project, using a per unit tax on capital invested in its region, at a rate τ_ℓ. After taxes are set, capital owners invest in the region(s) with the highest net return(s), central or regional governments raise their taxes, production takes place, and private consumption (net of mobility costs) occurs. When the project is completed, it generates a social benefit b for all residents of the region. We assume that $b < B$.[10]

In our environment, the federation is characterized by two dimensions of state capacity: *administrative* and *fiscal*.[11] The former measures the ability of regional bureaucracies to carry out projects in due time, and it is encapsulated by the probability π.[12] The latter gauges the capacity to collect local taxes. This dimension is modeled by assuming that local governments can only collect a fraction $\theta \in [0, 1]$ of their potential tax base, where θ measures the level of fiscal capacity.[13]

[7] To focus on the trade off between soft and hard budget constraints, we rule out spillovers across regions.

[8] Delays in local public works are prevalent both in developed and developing countries. See Guccio et al. (2014).

[9] Uniformity captures the idea that, for (non-modeled) constitutional reasons, the central government can neither set different tax rates contingent on which regional government asked for additional funds, nor make side-payments to any specific regional government. In the US, for example, federal taxes are required to be uniform across states by the "Uniformity Clause" (US Constitution, Article 1, Section 8, Clause 1).

[10] The difference between B and b reflects that some extra costs arise during the project's delay. For example, an incomplete park may affect pedestrian movement.

[11] See Hanson and Sigman (2013). Mann (1984) provides a general definition of state capacity as the infrastructural power of the state to enforce policy within its territory. Snyder (2001) and Ziblatt (2008) apply this concept to regional governments.

[12] Patil et al. (2013) document that public projects' delays in Indian states are mainly caused by administrative problems that arise during the land acquisition process.

[13] Arbetman-Rabinowitz et al. (2012) use a similar concept. Besley and Persson (2010), on the other hand, define

We summarize the timing in Figure 3.1. Terminal nodes represent the benefit of the project.

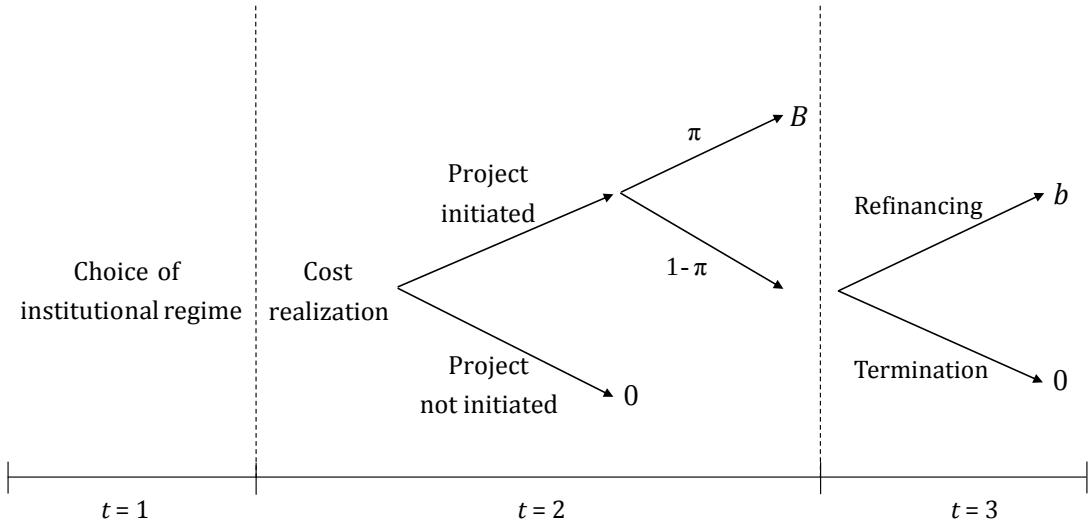


Figure 3.1: Timing.

3.2.3 DISCUSSION

Some features of the model deserve further comments. First, to focus on bailouts under partial decentralization, we assume that, under this institutional regime the central government does not intervene in a region to avoid its project's delay. Therefore, we do not need to incorporate into the model considerations of administrative capacity at the central level. In addition, regarding its fiscal capacity, we assume in the base model that the central government is fully efficient in this dimension. We relax this assumption in Section 3.6.

Second, we consider a discrete, regional public project instead of a continuous public good, as is common in most of the literature on tax competition.[14] Indivisibility fixes the type of competition between regions. As in Wildasin (1988), regions compete first in refinancing decisions, and then taxes are set accordingly in a context of imperfect capital mobility.[15] Moreover, this assumption combined with our specification of administrative competence is a simple way to analyze, via refinancing decisions, the interaction between levels of regional state capacity and different intergovernmental fiscal arrangements.

Third, regional governments and projects in the model are *ex ante* and interim identical. More precisely, all regional governments share the same exogenous levels of state capacity π and θ. Concerning projects, ex ante (i.e., in period 1) they are all characterized by the same configuration

fiscal capacity as an upper bound on the tax rate that a government can charge.

[14] To the best of our knowledge, the unique contribution to the local public finance literature that deals with discrete projects is Cremer et al. (1997).

[15] Akai and Sato (2008) and Köthenbürger (2011) also analyze models with this timing, but they consider income or wage taxation instead.

of exogenous social benefits b and B, and by the same probability density function $h(c)$. Interim (i.e., at the beginning of period 2), the cost c is realized and applies for all projects in all regions. But projects' outcomes can be different *ex post*. Indeed, at the end of period 2, some regions that have initially invested end up with complete projects, while others face incomplete ones.[16]

Also, in our model, regions do not have access to credit markets to refinance their incomplete projects. If regions can issue debt, they ultimately have to raise taxes to pay back their obligations. Hence, one can view our model as a convenient short cut capturing this feature of local public finances.

Finally, we assume that $c \leq b$. When $c > b$, under both institutional regimes incomplete projects are shut down. Therefore, we consider a parameter configuration of the model that isolates the case where the institutional comparison between partial and full decentralization is relevant.

3.3 FIRST BEST

In this section we analyze a benchmark for efficiency. Consider a social planner who makes all decisions, but cannot anticipate whether a project will be completed at the end of $t = 2$ (i.e., he has to carry out projects through the regional bureaucracies). We solve his decision problem backwards.

First, the refinancing decision in any region is independent of the planner's choice in any other region. This is because individual utilities are linear in income and the planner maximizes the sum of utilities. Thus, in any region, continuing an incomplete project is always optimal because $c \leq b$.

Moving back to the initial investment decision, the planner faces another separable problem between regions. Knowing the cost c, he initiates projects provided their expected benefit is higher than their expected cost (which includes a possible second round of financing). Let

$$c^*(\pi) \equiv \frac{\pi B + (1-\pi)b}{2-\pi}$$

denote the cost that makes the net expected regional welfare from initiating a project equal to zero. If $c \leq c^*(\pi)$, initial investment is efficient in any region ℓ; otherwise, not investing is the optimal decision.

For a given configuration of parameters (b, B), efficient investment decisions are depicted in Figure 3.2. A point in the (π, c) plane represents a project that costs c in a region with administrative capacity π.

When $0 \leq \pi \leq \pi^* \equiv b/B$, we have that $0 < c^*(\pi) \leq b$. Therefore, there exists a non-empty area NI^*, delimited by the thick curve $c^*(\pi)$, where it is optimal not to initiate projects. Below this curve, in the area I^*, it is efficient to undertake all projects. Ceteris paribus, as π increases

[16] An important feature of the model is risk neutrality. If individuals were risk averse, this could imply an insurance rationale for bailouts under partial decentralization.

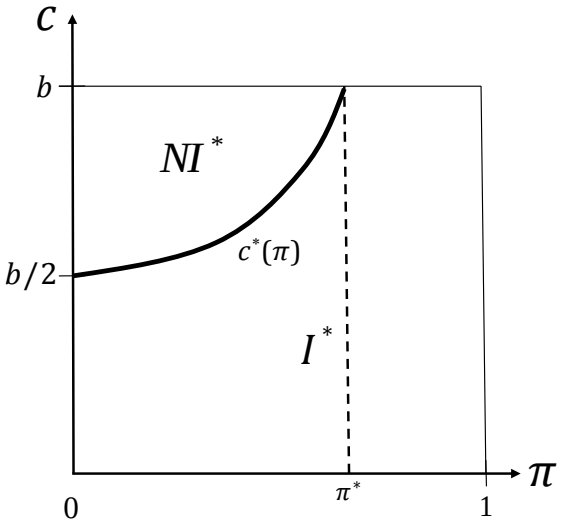


Figure 3.2: First Best.

the projects' net expected benefit increases. Thus $c^*(\pi)$ increases as well. When $\pi^* < \pi \leq 1$, $c^*(\pi) > b$: all projects are efficiently initiated.

3.4 PARTIAL DECENTRALIZATION

This section studies the partially decentralized regime. Given that $c \leq b$, under this regime incomplete projects are always refinanced by the central government using a uniform tax on capital. This implies that project initiation decisions of any given region may ultimately impact welfare of others, thus giving rise to a simultaneous game between regions in the second period (i.e., when the initial investment decision is made).

Region ℓ's expected welfare at the beginning of the second period is given by

$$\mathbb{E}W_\ell^{PD}(i_\ell, \mathbf{i}_{\boldsymbol{m}}) = \kappa(1-\tau^e) + \mathbb{1}_{\{i_\ell = I\}}[\pi B + (1-\pi)b - c], \tag{3.1}$$

where $\mathbf{i}_{\boldsymbol{m}}$ is the profile of investment decisions chosen by regions $m \neq \ell$, $\mathbb{1}_{\{i_\ell = I\}}$ is equal to 1 if region ℓ has initiated the project and to 0 otherwise, and τ^e is the expected tax.

The expected tax τ^e is obtained as follows. At the end of the second period, all projects' outcomes are realized. Let ω be a profile of outcomes, and denote by $\mathcal{B}(\omega)$ the number of completed projects in this particular realization of outcomes. For any profile ω, the central government mechanically sets a tax τ_ω to cover, at the beginning of the third period, the cost of refinancing $\sum_\ell \mathbb{1}_{\{i_\ell = I\}} - \mathcal{B}(\omega)$ incomplete projects. As this tax is uniform and exporting capital is costly, every household will invest in its own region. This implies that the tax base is $L\kappa$, and taxation is non distortionary. Hence, under the profile ω, the central government's budget

constraint is

$$\tau_\omega . L\kappa = \left[\sum_\ell \mathbb{1}_{\{i_\ell = I\}} - \mathcal{B}(\omega) \right] c.$$

Therefore, when deciding on initial investment before the outcome of projects are realized, each region faces the expected tax $\tau^e \equiv \mathbb{E}$, which satisfies

$$\tau^e . L\kappa = \left[\sum_\ell \mathbb{1}_{\{i_\ell = I\}} (1 - \pi) \right] c, \tag{3.2}$$

where the term in square brackets gives the expected number of bailouts. Substituting (3.2) into (3.1) and rearranging, we obtain

$$\mathbb{E}W_\ell^{PD}(i_\ell, \mathbf{i}_{\boldsymbol{m}}) = \kappa + \mathbb{1}_{\{i_\ell = I\}} \left[\pi B + (1 - \pi) \left(b - \frac{c}{L} \right) - c \right] - \sum_{m \neq \ell} \mathbb{1}_{\{i_m = I\}} (1 - \pi) \frac{c}{L}. \tag{3.3}$$

By inspection of (3.3), the effect of i_ℓ on $\mathbb{E}W_\ell^{PD}$ (captured by the term in square brackets) is independent of $\mathbf{i}_{\boldsymbol{m}}$. So, we can analyze the choice of i_ℓ just for a representative region ℓ.

Notice that each region only pays $1/L$ of the cost of refinancing its incomplete project, as this cost is shared through national taxation. Therefore, the central government's budget constraint generates a *common-pool fiscal externality*: any resident of ℓ is negatively affected by the possibility of an incomplete project in a region $m \neq \ell$. Let

$$c^{PD}(\pi) \equiv \frac{L[\pi B + (1 - \pi)b]}{L + 1 - \pi}$$

be the cost that makes the net expected regional welfare from initiating the project under partial decentralization equal to zero. The next proposition completely characterizes regional project initiation decisions under this institutional regime.

Proposition 3.4.1. *Consider the project initiation game under partial decentralization. Symmetric equilibria are as follows. If $c \leq c^{PD}(\pi)$, initial investment takes place in all regions. Otherwise, no region invests in equilibrium.*

Proof. See Appendix 3.9.1. □

In the Appendix we show that $c^{PD}(\pi) > c^*(\pi)$. Hence, we can establish the form of the inefficiencies that emerge under this institutional regime as follows.

Corollary 3.4.1. *Under partial decentralization, initial investments may occur in equilibrium when it is inefficient to do so.*

Inefficiencies involve *over investment*. This kind of inefficiency, driven by the common-pool fiscal externality, is well known (See Wildasin (1997) and Goodspeed (2002)). The following figure depicts equilibrium outcomes that emerge under partial decentralization.

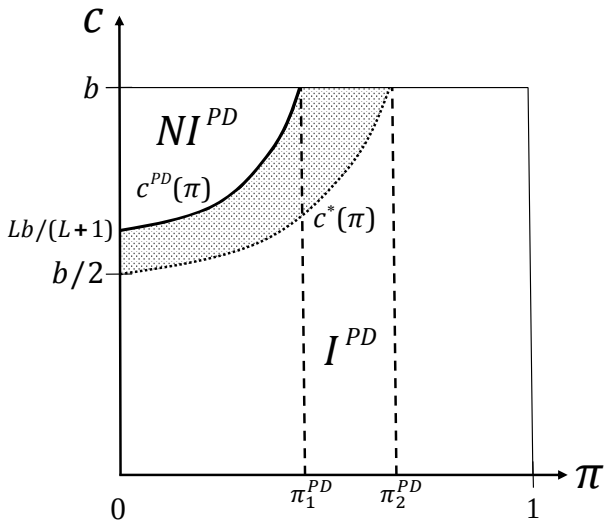


Figure 3.3: Partial Decentralization.

Analogous to the first best, when $0 \leq \pi < \pi_1^{PD} \equiv \frac{b}{L(B-b)+b}$, there exists a non-empty area (NI^{PD}) where projects are efficiently not initiated in any region, and another one (I^{PD}) where all projects are initiated. In the white area, regional decisions are optimal and so, in equilibrium, each region is expected to contribute an amount equal to the cost of refinancing by itself its incomplete project. Thus, interim expected welfare in each region coincides with the first best level. In the shaded area, when $c \in [c^*(\pi), c^{PD}(\pi)]$, the inefficient investment decision is adopted in equilibrium by all regions. When $\pi_1^{PD} \leq \pi \leq \pi_2^{PD} \equiv b/B$, $c^*(\pi) \leq b \leq c^{PD}(\pi)$, which implies that all projects are initiated in equilibrium. Projects in the shaded area are inefficiently initiated. Finally, when $\pi_2^{PD} < \pi \leq 1$, inefficient investments cannot emerge because the model is biased towards project initiation. Under these parameter conditions, partial decentralization replicates the first best outcome.

3.5 FULL DECENTRALIZATION

In this section we analyze the fully decentralized regime. In this case, a three-stage simultaneous game between regions emerges. First, regional governments take the initial investment decision. Second, the continuation decision is made. Finally, refinancing is achieved by levying taxes on capital employed in the region, in a context of tax competition.

But before solving the game backwards, it is convenient to describe how capital reacts to different tax rates. Given a profile of tax rates $\boldsymbol{\tau} = \{\tau_1, ..., \tau_L\}$ set by all regions, a household resident in region ℓ decides where to invest its capital endowment. Let $f_{\ell m}$ denote the amount of capital that this household invests in a region $m \neq \ell$, and let $\widetilde{M}$ be the set of regions $\widetilde{m} \neq \ell$ that have chosen the minimum tax rate $\widetilde{\tau}_\ell = \min\{\tau_m\}_{m \neq \ell}$. The following proposition characterizes the household's investment decision.

Proposition 3.5.1. *If* $\tau_\ell \geq \widetilde{\tau}_\ell$, $\sum_{\widetilde{m}\in\widetilde{M}} f_{\ell\widetilde{m}} = \tau_\ell - \widetilde{\tau}_\ell \geq 0$. *Otherwise,* $f_{\ell\widetilde{m}} = 0$.

Proof. See Appendix 3.9.2. □

As a household in region ℓ seeks to maximize net returns from its investments, its portfolio decision only depends upon the comparison between τ_ℓ and $\widetilde{\tau}_\ell$, and not between τ_ℓ and the whole profile of tax rates chosen by regions $m \neq \ell$. As expected, a household in region ℓ invests "abroad" in a region $\widetilde{m} \in \widetilde{M}$ provided $\tau_\ell > \widetilde{\tau}_\ell$. When only one region sets the minimum tax rate $\widetilde{\tau}_\ell$, it attracts all capital that leaves region ℓ. But if two or more regions choose the same tax rate $\widetilde{\tau}_\ell$, the value of the capital outflow $f_{\ell\widetilde{m}}$ that goes to each of these regions is undetermined. The intuition behind the expression in the proposition is straightforward, as capital leaves region ℓ until the marginal tax savings equal the marginal mobility cost. As expected, this flow increases with the difference $\tau_\ell - \widetilde{\tau}_\ell$.

When the regional tax rate τ_ℓ is lower or equal than $\widetilde{\tau}_\ell$, there is no capital outflow from region ℓ and its residents invest all their endowment "at home." Moreover, in this case, region ℓ receives capital inflows from other regions. But this does not benefit directly its residents because returns from these investments are consumed abroad, by residents in regions $m \neq \ell, \widetilde{m}$. Despite this fact, in the next section we show that these capital inflows have an important role in the determination of the equilibrium tax rates.

3.5.1 EQUILIBRIUM IN TAX RATES

In the remainder of the section, we solve the game between regions backwards. We start with the last stage, where regional governments set tax rates to refinance incomplete projects.

To obtain the equilibrium tax rates, in the Appendix we derive region ℓ's reaction function $\tau_\ell(\boldsymbol{\tau_m})$, where $\boldsymbol{\tau_m}$ denotes the profile of tax rates chosen by regions $m \neq \ell$. We have previously discussed that, depending upon the whole profile of tax rates, capital may leave or enter region ℓ from other regions. Despite these different possibilities, region ℓ's after-tax consumption monotonically decreases with τ_ℓ, and so does regional welfare. Hence, for any profile $\boldsymbol{\tau_m}$, the tax rate chosen in region ℓ should be the lowest tax rate that enables the regional government to raise c.

The reaction function $\tau_\ell(\boldsymbol{\tau_m})$ is built around the value $c/\theta\kappa$, which is the tax rate that a regional government would choose in autarky. When the profile $\boldsymbol{\tau_m}$ is such that $\tau_m < c/\theta\kappa$ for all m, region ℓ's optimal response is to tax strictly above the minimum tax $\widetilde{\tau}_\ell$. Despite the fact that this decision will trigger a capital outflow, this is the only way to ensure the project's refinancing. When all tax rates τ_m are set equal to $c/\theta\kappa$, region ℓ's optimal response is to replicate this level. Due to the way we model imperfect capital mobility, the tax collection's elasticity with respect to τ_ℓ is lower than one. This combined with the fact that region ℓ needs to collect enough revenues to refinance its incomplete project, makes tax undercutting not a profitable deviation. Finally, when the profile of tax rates is such that $\widetilde{\tau}_\ell > c/\theta\kappa$ or $\widetilde{\tau}_\ell = c/\theta\kappa$ but at least one region $n \neq \ell$ has chosen a tax rate $\tau_n > c/\theta\kappa$, region ℓ's optimal response is to tax strictly below $\widetilde{\tau}_\ell$. This decision

generates an inflow of capital that allows the government to raise sufficient revenues to pursue its incomplete project, thus moderating the tax burden on its residents.

Another important feature of region ℓ's reaction function is that it is non-continuous. Despite this fact, we can still characterize the Nash equilibria of this subgame as follows.

Proposition 3.5.2. *When all regions have decided to refinance their incomplete project, the unique symmetric Nash equilibrium in pure strategies is such that $\widehat{\tau}_\ell = c/\theta\kappa$. When there is at least one region that does not refinance, then all regions that refinance set*

$$\widehat{\tau}_\ell(0) \equiv \frac{1}{2}\left[\kappa - \sqrt{\kappa^2 - (4c/\theta)}\right].$$

Proof. See Appendix 3.9.3. □

Consider the tax competition subgame that emerges when all regions have decided to refinance their incomplete project. The equilibrium tax rate $\widehat{\tau}_\ell$ increases with the cost of the project c. Moreover, as in equilibrium nobody invests abroad, regions tax their own capital endowment without bearing any deadweight loss due to its mobility. Thus, the higher this endowment, the lower the equilibrium tax rate. Similarly, the higher the fiscal capacity θ, the lower the equilibrium tax $\widehat{\tau}_\ell$. In this case, region ℓ's welfare (net of the initial cost c) is $W_\ell^{FD} = \kappa + b - \frac{c}{\theta}$ in equilibrium. When $\theta < 1$, imperfect fiscal capacity implies that regions do not pay only the technical cost of completing the project c, but a higher, effective refinancing cost c/θ. The difference between these values corresponds to the cost of tax collection.

The proposition also shows that, when at least one region does not refinance (in which case it does not need to tax its population), region ℓ has to set the tax rate $\widehat{\tau}_\ell(0)$ to pursue its ongoing project. Thus, asymmetric taxation emerges as a possible equilibrium, as in Bucovetsky (1991) and Wilson (1991).[17] The tax rate $\widehat{\tau}_\ell(0)$ also decreases with the capital endowment κ and the fiscal capacity θ. With this tax rate, the resulting capital outflow is $\sum_{\widetilde{m}\in\widetilde{M}} f_{\ell\widetilde{m}} = \widehat{\tau}_\ell(0)$ and the equilibrium regional welfare (net of the initial cost c) is

$$\begin{aligned} W_\ell^{FD} &= \left(\kappa - \textstyle\sum_{\widetilde{m}\in\widetilde{M}} f_{\ell\widetilde{m}}\right)\left(1 - \widehat{\tau}_\ell(0)\right) + \textstyle\sum_{\widetilde{m}\in\widetilde{M}} f_{\ell\widetilde{m}} - \frac{1}{2}\left(\textstyle\sum_{\widetilde{m}\in\widetilde{M}} f_{\ell\widetilde{m}}\right)^2 + b \\ &= \kappa + b - T(c,\theta), \end{aligned}$$

where

$$T(c,\theta) = \frac{c}{\theta} + \frac{[\widehat{\tau}_\ell(0)]^2}{2}$$

measures the total refinancing cost, which comprises the effective refinancing cost c/θ, plus the deadweight loss $[\widehat{\tau}_\ell(0)]^2/2$ of financing the continuation of the project through a distortionary tax. This distortion is due to mobility costs incurred by owners of capital seeking to avoid taxation in

[17] The main difference with their result is that variations in tax rates across regions are not originated from an ex ante regional asymmetry, but rather from the possibility that some regions may end up with incomplete projects ex post.

region ℓ. Therefore, distortionary taxation only emerges in some final nodes of the tax competition subgame, when at least one region does not refinance. More importantly, the likelihood of these final nodes depends upon the level of regional administrative capacity π.

3.5.2 REFINANCING

At the beginning of the second stage, the strategy for region ℓ is $r_\ell \in \{R, NR\}$, where R (NR) denotes "refinancing" ("not refinancing"). Conditional on $\mathbf{i} = (i_1, ..., i_L)$, and given $\mathbf{r_m}$, the profile of refinancing decisions chosen by regions $m \neq \ell$, region ℓ's welfare is

$$W_\ell^{FD}(r_\ell = R, \mathbf{r_m}, \mathbf{i}) = \kappa + \mathbb{1}^\ell_{\{I,inc\}}\{[b - \mathbb{1}^m_{\{I,inc,R\}}\frac{c}{\theta} - (1 - \mathbb{1}^m_{\{I,inc,R\}})T(c,\theta)] - c\},$$

and

$$W_\ell^{FD}(r_\ell = NR, \mathbf{r_m}, \mathbf{i}) = \kappa - \mathbb{1}^\ell_{\{I,inc\}}c$$

where $\mathbb{1}^\ell_{\{I,inc\}}$ is equal to 1 if region ℓ initiated a project that has remained incomplete at $t = 2$, and to 0 otherwise, and $\mathbb{1}^m_{\{I,inc,R\}}$ is equal to 1 if all regions $m \neq \ell$ initiated their project, have not completed it in due time but decided to refinance them in $t = 3$, and to 0 otherwise.

The following proposition characterizes the Nash equilibria of this subgame.

Proposition 3.5.3. *Let c_1 denote the value of c that makes the total refinancing cost $T(c,\theta)$ equal to the benefit b, and $c_2 > c_1$ the value of c that makes the effective refinancing cost c/θ equal to the benefit b. When all regions face an incomplete project, they refinance them in equilibrium provided $c \leq c_2$. Otherwise, no region refinances. When there is at least one region that does not need refinancing, region ℓ refinances its incomplete project provided $c \leq c_1$.*

Proof. See Appendix 3.9.4. □

In the Appendix we show that $c \leq c_1$ ($c \leq c_2$) implies $T(c,\theta) \leq b$ ($c/\theta \leq b$). When all regions failed to complete their project in due time, the Nash equilibria of this refinancing subgame depend upon the cost c. When $c < c_1$, refinancing is a dominant strategy. But when $c_1 \leq c \leq c_2$, a standard coordination game emerges, with two Nash equilibria. The first one is trivial: despite the fact that $c \geq c_1$, when all regions decide to refinance, these strategies form a Nash equilibrium because, as there will be no capital flows (and thus no deadweight loss due to distortionary taxation), only the fact that $c/\theta \leq b$ matters. But if at least one region has decided not to refinance, the other regions also shut down their project because $c \geq c_1$. In this case, these regions face an endogenous hard budget constraint due to other regions' decisions, as in Qian and Roland (1998). We focus on the Nash equilibrium in which all regions refinance because it is strong (Aumann (1959)) and coalition-proof (Bernheim et al. (1987)). Finally, when $c_2 \leq c$, not refinancing is a dominant strategy. Regions face again an endogenous hard budget constraint, but this time as a consequence of their imperfect regional fiscal capacity. As $c \leq b$, this outcome is inefficient.

In all other subgames, when there is at least one region that does not need refinancing because it executed its project in due time, region ℓ refinances its incomplete project provided $c \leq c_1$. If $c > c_1$, the total cost from refinancing is higher than the benefit b, pushing region ℓ to shutdown its incomplete project. Once more, this is an inefficient outcome.

3.5.3 PROJECT INITIATION

Anticipating refinancing equilibria, regional governments simultaneously adopt the project initiation decision. The next proposition characterizes the Nash equilibria of the investment stage under this regime.

Proposition 3.5.4. *Consider the project initiation game under full decentralization. Let $c_R^{FD}(\pi)$, $c_{NAR}^{FD}(\pi)$ and $c_{NR}^{FD}(\pi)$ be the costs that make the net expected regional welfare equal to zero (i) when incomplete projects are refinanced regardless of the continuation decisions of other regions, (ii) when incomplete projects are refinanced only if all other regions do so, and (iii) when incomplete projects are never refinanced, respectively. There exists thresholds π_1^{FD}, π_2^{FD} and π_3^{FD} such that the symmetric Nash equilibria are as follows:*

1. *When $0 \leq \pi \leq \pi_1^{FD}$, initial investment takes place in all regions provided $c \leq c_R^{FD}(\pi)$. Otherwise, no region invests in equilibrium.*
2. *When $\pi_1^{FD} < \pi \leq \pi_2^{FD}$, initial investment takes place in all regions provided $c \leq c_{NAR}^{FD}(\pi)$. Otherwise, no region invests in equilibrium.*
3. *When $\pi_2^{FD} < \pi \leq \pi_3^{FD}$, initial investment takes place in all regions provided $c \leq c_{NR}^{FD}(\pi)$. Otherwise, no region invests in equilibrium.*
4. *When $\pi_3^{FD} < \pi \leq 1$, initial investment takes place in all regions.*

Proof. See Appendix 3.9.5. □

In the Appendix we characterize the probability thresholds π_1^{FD}, π_2^{FD} and π_3^{FD}, and we show that cost thresholds $c_R^{FD}(\pi)$, $c_{NAR}^{FD}(\pi)$ and $c_{NR}^{FD}(\pi)$ are lower than $c^*(\pi)$. Hence, we can establish the types of inefficiencies that emerge under this institutional regime, as follows.

Corollary 3.5.1. *Under full decentralization, equilibrium outcomes can be inefficient for three reasons: (i) initial investments do not take place in equilibrium when it is efficient to do so, (ii) initial investments take place in equilibrium and are efficient, but incomplete projects are not refinanced when it is efficient to do so, or (iii) initial investments take place in equilibrium and are efficient, but incomplete projects are refinanced using distortionary capital taxes.*

The following figure depicts equilibrium outcomes that emerge under full decentralization.

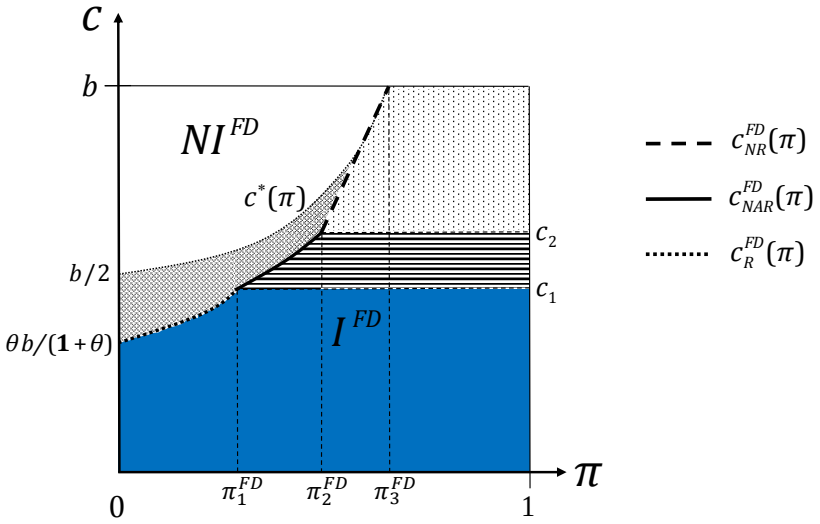


Figure 3.4: Full Decentralization.

In the non-empty area NI^{FD}, delimited from below by the thick curves representing the cost thresholds $c_R^{FD}(\pi), c_{NAR}^{FD}(\pi)$ and $c_{NR}^{FD}(\pi)$, projects are not initiated. In the complementary area, denoted by I^{FD}, all projects are initiated.

When $c \in [c^*(\pi), b]$, efficient decisions are adopted. In all other areas, either the projects' initiation and continuation decisions are distorted or the region refinances bearing deadweight losses. When $0 \leq \pi \leq \pi_1^{FD}$, two different types of inefficiency emerge. First, in the darker area when $c \in [0, c_R^{FD}(\pi)]$, initiation and continuation decisions are optimal, but refinancing is done bearing deadweight losses generated by imperfect regional fiscal capacity or distortionary capital taxation. Second, as $c_R^{FD}(\pi) < c^*(\pi)$, the condition for project initiation is stricter with full decentralization than for the social planner. Therefore, in the shaded area when $c \in [c_R^{FD}(\pi), c^*(\pi)]$, investments are not initiated in equilibrium, despite the fact that they are efficient. Underinvestment is due to i) distortionary refinancing, when $c \in [c_R^{FD}(\pi), c_1]$, and ii) endogenous hard budget constraints, when $c \in [c_1, c^*(\pi)]$.[18]

When $\pi_1^{FD} < \pi \leq \pi_2^{FD}$, a new type of inefficiency emerges. In the dashed area, projects are initiated but only refinanced when all regions do so. Again, there is underinvestment when $c \in [c_{NAR}^{FD}(\pi), c^*(\pi)]$, but only due to endogenous hard budget constraints.

As $\pi_2^{FD} < \pi \leq \pi_3^{FD}$, the last type of inefficiency emerges. In the dotted area projects are initiated but never refinanced. There is still a shaded area where, because of endogenous hard budget constraints, projects are not initiated.[19]

Finally, when $\pi \geq \pi_3^{FD}$, the model is biased towards project initiation. Incomplete projects are

[18]The former is the analogue of the Zodrow and Mieszkowski (1986) result, while the latter has been analyzed by Besfamille and Lockwood (2008), in a setting where an exogenous hard budget constraint is imposed to regional governments.

[19]Figure 3.4 depicts full decentralization outcomes when $\theta < 1$. If $\theta = 1$, $c_2 = b$ and $\pi_2^{FD} = \pi_3^{FD}$: the area where projects are never refinanced vanishes.

either refinanced in a distortionary way, shut down in some terminal nodes of the tax competition subgame, or never finished.

3.6 OPTIMAL INSTITUTIONAL REGIME

3.6.1 MAIN RESULTS

In the initial period, there is an institutional choice between partial and full decentralization. At this stage, the Congress observes projects' benefits b, B, the regional capital endowment κ and state capacities (π, θ), and knows that the cost c is distributed according to a strictly positive probability density function $h(c)$ on $[0, b]$. The Congress chooses the optimal regime by maximizing

$$\mathbb{E}W^{IR} = \int_0^b \mathbb{E}W_\ell^{IR} h(c) dc,$$

where $\mathbb{E}W^{IR}$ is the expected welfare of a representative region, under the institutional regime $IR \in \{PD, FD\}$.[20]

The main goal of this paper is to evaluate how the regime choice is affected by the level of regional state capacity. We first characterize the relationship between $\mathbb{E}W^{IR}$ and the pair (π, θ). Taking into account equilibrium decisions and outcomes, we show that $\mathbb{E}W^{FD}$ is a continuous, everywhere differentiable, increasing, convex function of the administrative capacity π. Moreover, it also increases with the fiscal capacity θ. Therefore, under full decentralization, an increase in either π or θ increases $\mathbb{E}W^{FD}$, as suggested by the decentralization literature.[21] But this assertion does not necessarily imply that full decentralization dominates for high levels of state capacity. The reason is that $\mathbb{E}W^{PD}$ is a continuous, increasing, convex function of the administrative capacity π.[22] Hence, the comparison between both regimes is not a priori evident.

To make progress, we introduce the following assumption which ensures that the intersection between $\mathbb{E}W^{FD}$ and $\mathbb{E}W^{PD}$ is unique.

Assumption 3.6.1. *Assume that:*

1. *The cost c is distributed uniformly on $[0, b]$.*
2. *The benefit b satisfies $B/2 \leq b < B$.*
3. *The number of regions L satisfies $\underline{L} \leq L \leq \overline{L}$.*[23]

[20] In a one-shot version of the model, the Congress could wait until the realization of the cost to choose the optimal interim regime. But in a repeated version of the model, it seems realistic to assume that changing the institutional regime after each realization of c would be too costly, which justifies our focus on the optimal ex-ante regime.

[21] See, among others, Bird (1995), Litvack et al. (1998) and Loayza et al. (2014).

[22] See the Online Appendix.

[23] The definitions of $\underline{L}$ and $\overline{L}$ are given in the Appendix.

The following proposition characterizes the optimal choice between partial and full decentralization.

Proposition 3.6.1. *Suppose Assumption 3.6.1 holds and let* $\theta_0 \equiv 2L/(1+L^2)$ *be the regional fiscal capacity that makes* $\mathbb{E}W^{PD}$ *equal to* $\mathbb{E}W^{FD}$ *when* $\pi = 0$*. Then:*

1. *When* $\theta < \theta_0$*, partial decentralization dominates for all values of* π*.*
2. *When* $\theta \geq \theta_0$*, there exists a unique threshold* $\hat{\pi}(\theta)$ *such that, when* $\pi \leq \hat{\pi}(\theta)$*, full decentralization dominates. Otherwise, partial decentralization dominates.*
3. *When* $\pi = 1$*, both regimes are efficient.*

Proof. See Appendix 3.9.6. □

Figure 3.5 illustrates this result. Each point in the (θ, π) plane represents the regional state capacity that prevails in the federation. In the Appendix, we show that $\hat{\pi}(\theta)$ increases with θ.[24]

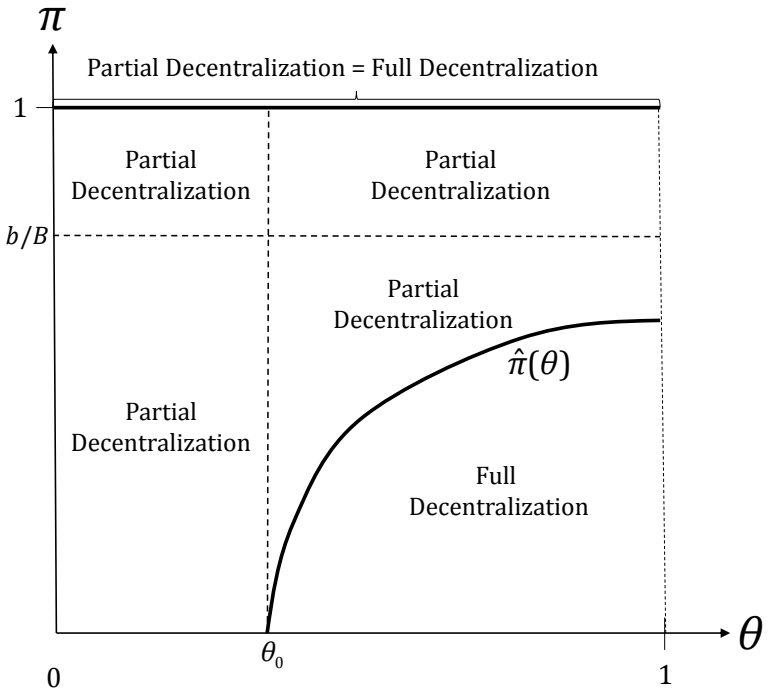


Figure 3.5: Partial Vs. Full Decentralization.

When regional fiscal capacity satisfies $\theta < \theta_0$, refinancing incomplete projects under full decentralization is too costly. Therefore, regardless of the level of regional administrative capacity, partial decentralization dominates.

[24]When Assumption A does not hold, the model becomes analytically untractable and thus unicity of $\hat{\pi}(\theta)$ cannot be ensured. In order to verify whether our results would hold under parameter configurations of the model that do not satisfy those conditions, we have simulated the model and replicated Figure 3.5. All simulations confirmed the results presented in Proposition 3.6.1, and are available upon request.

As expected, full decentralization may dominate when the regional fiscal capacity is relatively high. But interestingly, this occurs for relatively low levels of the regional administrative capacity, i.e., when $\pi \leq \hat{\pi}(\theta)$. The intuition for this result is the following. Under full decentralization, tax distortions emerge if at least one region does not need refinancing. But when π is low, the likelihood of this event is also low. Thus, distortions under full decentralization cost less than inefficient investments made under partial decentralization, where investments need to be refinanced with a relatively high probability. On the other hand, when π is relatively high, the expected welfare cost of inefficiently initiated projects under partial decentralization is low (or even zero) because it is very likely that they will be completed at the end of $t = 2$.

When π is above b/B, the model is biased towards project initiation and under partial decentralization, outcomes are efficient. On the other hand, under full decentralization, the likelihood of facing capital mobility costs or shutting down the project is still positive. Thus, partial decentralization dominates. But if π increases further and converges to one, the likelihood and welfare cost of full decentralization's distortions decrease, attenuating partial decentralization's dominance. In the limit, when π is equal to one, both regimes yield optimal outcomes.

These results clarify how the level of regional state capacity prevalent in a federation affects the trade-off between partial and full decentralization. On the one hand, our results confirm that a high level of fiscal capacity is a necessary condition for full decentralization to be the optimal institutional regime. But, we also caution against the position held by some authors suggesting that high levels of regional state capacity are necessary for successful decentralization reforms. First, our model shows that high levels of regional administrative or fiscal capacity do not always imply that full decentralization should dominate. Second, it is still plausible that full decentralization dominates even for low levels of administrative capacity.

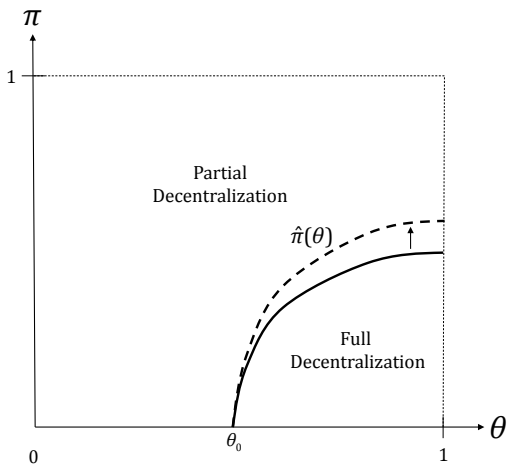
3.6.2 COMPARATIVE STATICS

In this subsection we analyze how changes in the key parameters of the model affect the comparison between partial and full decentralization.[25]

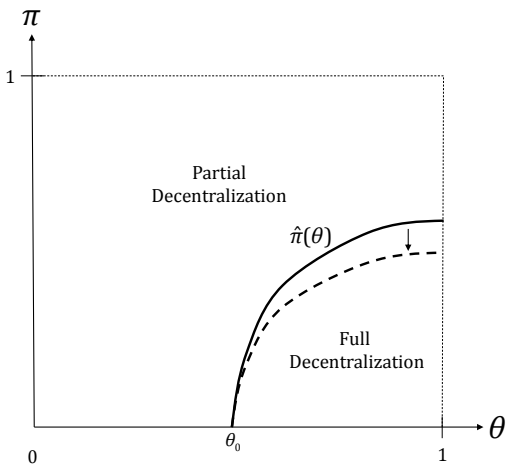
Change in the capital endowment κ. Figure 3.6(a) shows that an increase in κ favors full decentralization, in the sense that this regime dominates in a larger area of the plane (θ, π). The intuition of this result hinges on two facts. First, the cost of distortions generated under partial decentralization do not depend upon the value of κ. Second, an increase in the regional stock of capital increases expected welfare under full decentralization, because as the tax $\widehat{\tau}_\ell(0)$ decreases with κ, and so does the mobility cost of capital.

Change in the project's benefit B. Figure 3.6(b) shows that an increase in B always favors partial decentralization. The reason is that when B increases, the welfare loss due to overinvestment under

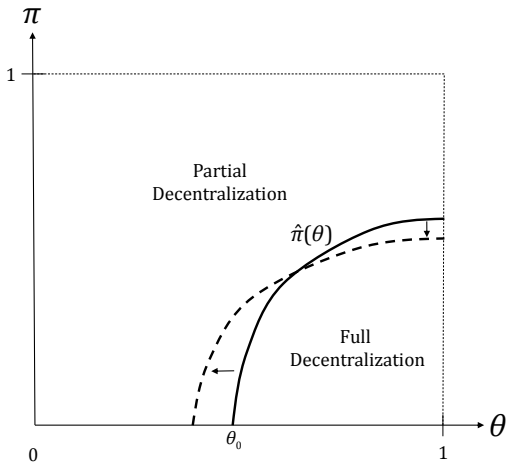
[25] Proofs corresponding to this section are contained in the Online Appendix.



(a) Increase in Capital Endowment κ.



(b) Increase in Benefit B.



(c) Increase in Number of Regions L.

Figure 3.6: Comparative Statics.

partial decentralization increases less than the corresponding welfare loss due to underinvestment under full decentralization.

Change in the number of regions L. Figure 3.6(c) shows that an increase in L has two effects. On the one hand, for relatively high levels of the regional fiscal capacity, partial decentralization dominates in an area where full decentralization used to be the optimal regime, and vice versa for low levels of regional administrative capacity. The intuition for this result hinges on the relative costs of an increase in L. On the one hand, the common-pool fiscal externality under partial decentralization gets larger, and thus more projects are inefficiently initiated. On the other hand, under full decentralization, the likelihood that a given region has to bear deadweight losses due to capital mobility or to its project's shutdown increases as well. When the administrative capacity π is relatively low, the cost of the latter is lower; the opposite holds for relatively high values of π.

3.6.3 AN EXTENSION: DISTORTIONARY NATIONAL TAXATION

So far, the tax set by the central government to bail out regions under partial decentralization was non distortionary. This is arguably a strong assumption: typically national taxes also generate deadweight losses, so assuming non-distortionary national taxation underestimates the cost of the common-pool fiscal externality. To address this issue, in this section we model distortions in national taxation by assuming that the cost of bailing out an incomplete project under partial decentralization is $c + \lambda$, where $0 < \lambda < b$.[26] Although this change does not affect the fully decentralized regime, outcomes in the other regime are modified as shown by the following figure.[27]

As expected, the central government refinances fewer projects than before. Indeed, only when $c \leq b - \lambda < b$, incomplete projects are bailed out; otherwise, the central government faces an endogenous hard budget constraint. When $0 \leq \pi \leq \pi_1^{PD,\lambda} \equiv \pi_1^{PD} - \frac{\lambda L}{L(B-b)+b}$, despite the fact that the central government bails out all projects with costs lower than $b - \lambda$, only those for which

$$c \leq c^{PD,\lambda}(\pi) \equiv c^{PD}(\pi) - \frac{(1-\pi)\lambda}{L+1-\pi}$$

are initiated and refinanced. Otherwise, projects are not undertaken because refinancing them is too costly. Projects falling within the shaded area are inefficiently initiated. But note that these are necessarily fewer than in the case in which national taxation is non distortionary because $c^{PD,\lambda}(\pi) < c^{PD}(\pi)$. As π increases, the cost interval where these inefficient projects are initiated when the central government refinances vanishes. Thus, when the central government does not bail out regions, no project is undertaken in equilibrium, as shown when $\pi_1^{PD,\lambda} <$

[26] This assumption is a shortcut that captures, among other things, cost differences in complying with central and regional tax systems. As stressed by Slemrod and Venkatesh (2002), these differences are substantial: nearly 70% of their compliance spending was devoted by firms to federal government's compliance, whereas almost 25% was spent on regional and local compliance. If, for the sake of simplicity, we normalize regional compliance costs to zero, λ measures the cost difference in complying with the central government.

[27] Proofs corresponding to this section are contained in the Online Appendix.

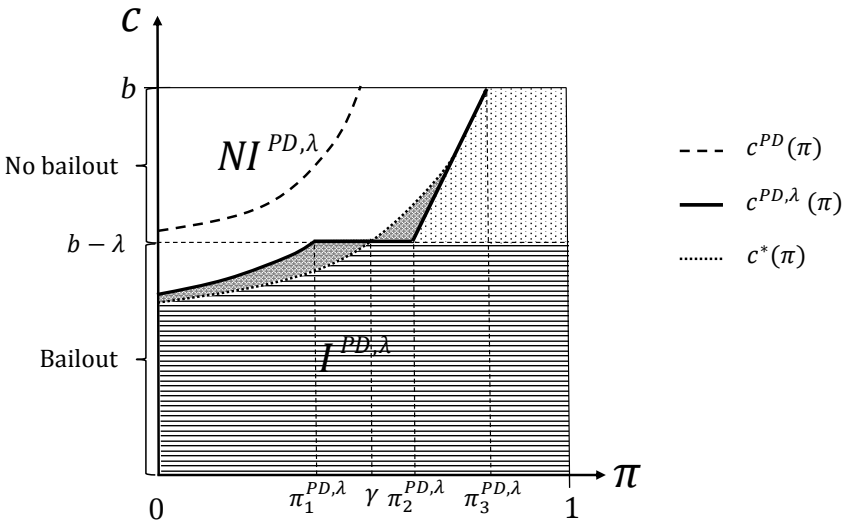


Figure 3.7: Partial Decentralization under Distortionary National Taxation.

$\pi \leq \pi_2^{PD,\lambda} \equiv \pi_2^{PD} - \frac{\lambda}{B}$. Note that, for $\pi \in [\gamma, \pi_2^{PD,\lambda}]$ we have $c^{PD,\lambda}(\pi) < c^*(\pi)$. This implies the opposite type of distortion: in the shaded area, regions inefficiently underinvest. Then, as $\pi_2^{PD,\lambda} \leq \pi \leq \pi_3^{PD,\lambda} \equiv b/B$, all projects that would be refinanced are initiated. But this does not hold for projects where $c > b - \lambda$. These projects are undertaken provided $c \leq \pi B$; otherwise they are not initiated, and again there is inefficient underinvestment in the shaded area. Finally, when $\pi_3^{PD,\lambda} \leq \pi \leq 1$, all projects are initiated, despite the fact that some with high cost would not be refinanced if they remain incomplete in the second period.

The following proposition compares this regime with full decentralization.

Proposition 3.6.2. *When the value of the deadweight loss λ increases, full decentralization dominates in a larger area of the plane (θ, π).*

Proof. See Online Appendix. □

Clearly, increasing λ favors full decentralization. Although expected, the intuition of this result should not be based on the mere assertion that by increasing the cost of bailouts, expected welfare under partial decentralization should automatically decrease. In fact, when λ increases, this institutional regime does generate more costly bailouts than before, inefficient project's shutdown and underinvestment. But when the administrative capacity is relatively low, a higher λ also reduces the size of the cost interval where projects are inefficiently initiated. Despite these countervailing impacts, we can show that the negative effects mentioned previously dominate, and thus expected welfare under this regime decreases with the deadweight loss λ.

3.6.4 ILLUSTRATION USING INTERNATIONAL DATA

Our model is normative and admittedly stylized. However, it is still interesting to compare our results to how fiscal decentralization varies with state capacity in the data.

To conduct such comparison we merge three data sets containing information for various developed and developing countries at an annual frequency. In the first place, we proxy fiscal capacity using the "Political Extraction Index" elaborated by Arbetman-Rabinowitz et al. (2012). The index essentially estimates the actual to the potential share of taxes to GDP.[28] Second, we summarize administrative capacity using the "Government Effectiveness Index" estimated by the World Bank.[29] [30] Finally, we capture the degree of fiscal decentralization in each country based on each country's *fiscal gap*, i.e. the share of subnational revenues not covered by subnational tax revenues. This time we use the database compiled by Gadenne and Singhal (2014), originally taken from the Government Finance Statistics of the International Monetary Found.

Our resulting dataset covers 45 countries and twelve years in the 1996-2010 period.[31] We average each country's measures of fiscal capacity, administrative capacity, and fiscal gap across time. A country is classified as "less" ("more") decentralized, if its fiscal gap is below (above) the median across countries. Figure 3.8 summarizes the data, where dashed lines depict the median level of state capacity in each dimension.[32]

Two observations stand out from Figure 3.6(a). First, on the north of the graph there is no clear pattern of fiscal decentralization. This fact resembles the normative prescription in our model, according to which federations should be indifferent between partial and full decentralization for very high levels of administrative capacity. Second, on the southeastern region the majority of countries exhibit high degrees of decentralization. The counterpart in our model is evident, given that full decentralization dominates when administrative capacity low and fiscal capacity is sufficiently high.

3.7 RELATED LITERATURE

This paper is related to various strands of the literature. First, other contributions have analyzed the trade-off between partial and full decentralization. Brueckner (2008) presents a Tiebout-type model, where local governments exert effort to reduce the cost of a local public good, private developers build houses and heterogeneous consumers decide on their location. Under partial decentralization, the federal government taxes the population and transfers the tax collection to jurisdictions, on an equally per capita basis. Then, local governments choose the level of effort

[28] The data is available from http://thedata.harvard.edu/dvn/dv/rpc.

[29] See http://info.worldbank.org/governance/wgi/index.aspx#home.

[30] The latter statistics have already been used to measure fiscal and administrative capacities, respectively, in previous studies. See Hanson and Sigman (2013) for a survey.

[31] Specifically, we have observations for the years 1996, 1998, 2000, 2002 and 2003 to 2010. We only keep countries with data for at least ten years.

[32] Data of state capacity is national, rather than regional. However, given that our model has symmetric regions, we can assimilate regional state capacities as being the level of state ability that prevails in the federation.

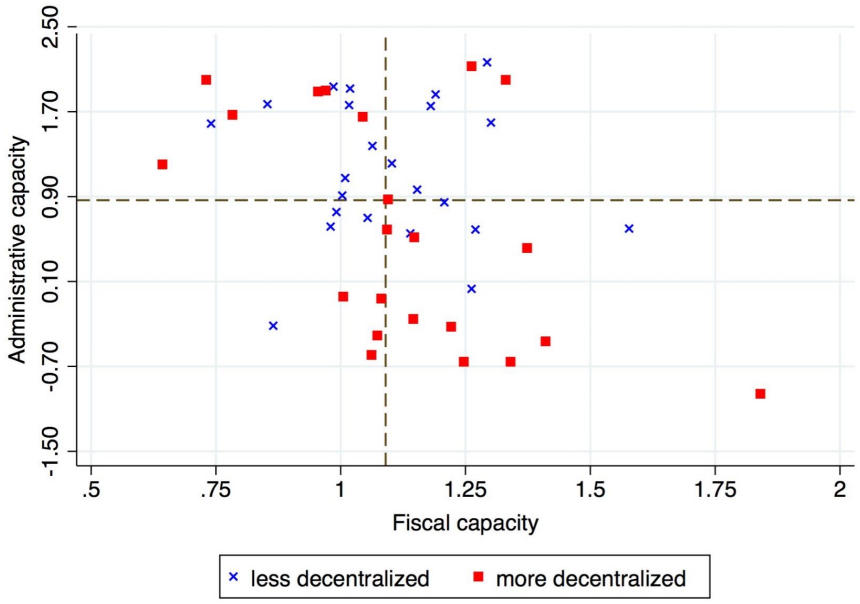


Figure 3.8: State Capacity and Decentralization in International Data.

and local public good provision, and finally Tiebout sorting occurs. Under full decentralization, the unique difference is that each local government sets a non-distortionary property tax to cover the public good's cost. Brueckner (2008) finds that full decentralization always dominates, because the uniformity of the transfer under partial decentralization generates less variety of local public goods, and thus a worse preference matching. Partial decentralization can be optimal, provided local governments are of a Leviathan type. Here, we do not find a complete regime dominance, even if regional governments are benevolent. Peralta (2011) compares both regimes, focusing on accountability issues at the local level. She presents a political economy model, with benevolent and self-interested, rent-seeking local politicians, who tax their jurisdiction and provide local public goods, in a context of double asymmetric information because voters observe neither the politician's type nor the local public good's cost. Two equilibria emerge: a pooling equilibrium, when the rent-seeker politician mimics a benevolent one; and a separating equilibrium, when he extracts rents, and thus he is replaced in the following election. Peralta (2011) shows that partial decentralization improves politician's selection (i.e., voting out rent-seekers) whereas full decentralization fosters discipline (i.e., giving incentives to rent seekers to behave as benevolents). This last regime dominates when the proportion of rent seekers is low. Other authors have adopted different definitions of partial decentralization. Janeba and Wilson (2011) and Hatfield and Padró i Miquel (2012) define partial decentralization when a subset of public goods are exclusively funded and provided by local governments. These authors find that devolving some public goods to local authorities is always optimal. Janeba and Wilson (2011) trade-off inefficient provision decided by a minimum winning coalition at the central level against distortionary taxation and low public good provision in a context of capital tax competition under full decentralization. In a voting model,

Hatfield and Padró i Miquel (2012) obtain that devolution serves as a commitment device against excessive capital taxation chosen when individuals vote on the central provision of public goods. Joanis (2014) defines partial decentralization as "shared responsibility", an institutional regime where both the central and the regional government participate in the funding of a given public good. In his model, central and local, rent-seeking, politicians simultaneously decide how much fiscal resources invest in the provision of this public good. Despite their preferences for rents, they also invest in such provision because their aim is to manipulate their reelection probability, in a context of weak accountability, where the electorate is unable to assess the contribution of each level of government to the public good provision. Partial decentralization obtains, balancing loses in productive complementarities between both levels of government and the lack of accountability at the local level. The main differences between our paper and these contributions hinges on the fact that we study a different trade-off between partial and full decentralization, namely inefficient bailouts and projects overprovision vs. capital tax competition and projects underprovision. Moreover, these articles do not incorporate regional state capacity into the analysis, and thus they do not consider it as a relevant factor affecting the trade-off.

The paper is also related to an important set of contributions that analyze the pros and cons of different types of regional budget constraints in federations. On the one hand, the optimality of hard budget constraints has been studied by Qian and Roland (1998) and Inman (2003); whereas the possibility that they may be inefficient has been raised by Besfamille and Lockwood (2008). The main differences between Besfamille and Lockwood (2008) and this paper are the following. First, they do not describe the institutional regime that hardens regional government's budget constraints; they simply assume that the federal government is able to impose, exogenously and at no cost, a hard budget constraint to local governments. Here, we analyze full decentralization as the institutional regime that hardens regional governments' budget constraint. Second, the authors compare, from a normative point of view, soft and hard budget constraints at the interim stage (in other words, project by project) whereas we take an ex ante perspective, more suited to an institutional comparison. Finally, they do not consider how different levels of regional administrative capacity affect the trade-off between partial and full decentralization, which is one of our main concerns. On the other hand, Wildasin (1997), Goodspeed (2002), **?** and Crivelli and Staal (2013) describe how bailouts in federations distort, via a common-pool fiscal externality, decisions at the regional level. Silva and Caplan (1997), Caplan et al. (2000) and Köthenbürger (2004) claim that, under some conditions, a regime with decentralized leadership, where the central government sets intergovernmental transfers after regional governments have adopted their own policy, may give a more efficient outcome than a regime with hard budget constraints. This result relies on a second best argument, and thus needs some pre-existing distortion in the form of public goods or tax spillovers to hold. Finally, Sanguinetti and Tommasi (2004) analyze the trade off between hard and soft budget constraints in the "rules vs. discretion" tradition. In their model, regions derive utility from the consumption of a private good and a national public good. Each region is endowed with an exogenous random level of output. The federal government has

an exogenous level of resources to finance the national public good and to set transfers to the regions. Two regimes are considered. In the first, the federal government commits ex ante to a transfer to each region which cannot be conditioned on the income shock (possibly because of incomplete information), and thus offers no insurance. In the second, the federal government fully accommodates to the requests for transfers that each region makes. This regime has the advantage of offering full insurance, but at the cost of a common pool problem where the national public good is underprovided. The authors find the conditions for which either one of the regimes dominate.

Finally, the paper is related to a recent literature that studies empirically how, in contexts of decentralized regimes, local state capacity impacts public outcomes. Steiner (2010) measures local governments' capacity using an index of resources available to local governments, and another one that captures the level of technical and administrative capacity of district governments. She finds evidence that both household consumption and school enrollment are positively related with the level of capacity of district governments. Loayza et al. (2014) evaluate how budget size and allocation process, local capacity, local needs, and political economy considerations (four factors that are usually considered important features that affect the effectiveness of decentralization reforms) affect municipal budget execution rate in Peru. The authors find that budget size and local capacity are the statistically most important constraints that explain municipal budget execution rates. Bandyopadhyay and Green (2012) show that higher percentage of residents from centralized ethnic groups imply higher development indicators at the local level in Uganda. Finally, Acemoğlu et al. (2014) study the impact of municipal state capacity on public goods provision in Colombia. An important feature of their paper is the consideration of spillovers: when a municipality invests in its state capacity, it also generates positive effects on neighboring municipalities. They empirically confirm that state capacity decisions are indeed strategic between municipalities, and with large effects on local prosperity. All these papers take the intergovernmental institutional setting as given, and thus do not analyze different regimes, as we do.

3.8 CONCLUSION

This paper presents a model featuring a central government and regional authorities. The latter are characterized their levels of administrative and fiscal capacities. We analyze two fiscal regimes. Under partial decentralization, regional governments rely on central bailouts to refinance previously started projects. Hence, regions face soft budget constraints and can overinvest in local public projects. Under full decentralization, regional governments cannot rely on central bailouts and face hard budget constraints. In this scenario, capital tax competition increases the marginal cost of public funds and regional governments may underinvest.

The main goal of the paper is to conduct a normative comparison between these regimes and determine how different levels of regional state capacity affect this comparison. As expected, when the regional fiscal capacity is low, partial decentralization dominates. But contrary to the

common wisdom, we find that full decentralization may be optimal even when the regional administrative capacity is low.

An interesting route for further research is to incorporate ex ante asymmetries between regions (either in capital endowments or in state capacities). This would yield a more suitable framework to deliver quantitative assessments, and would be an essential feature to endogenize regional state capacity formation under different fiscal regimes.

3.9 APPENDIX

3.9.1 PROOF OF PROPOSITION 3.4.1

The government of region ℓ anticipates that its net expected welfare from investing in the project is

$$\kappa + \mathbb{1}_{\{i_\ell = I\}} \left[\pi B + (1-\pi)\left(b - \frac{c}{L}\right) - c \right] - \sum_{m \neq \ell} \mathbb{1}_{\{i_m = I\}} (1-\pi) \frac{c}{L}, \tag{3.4}$$

whereas its net expected welfare from not investing is

$$\kappa - \sum_{m \neq \ell} \mathbb{1}_{\{i_m = I\}} (1-\pi) \frac{c}{L}.$$

So, for any region ℓ, initiating the project is a dominant strategy if

$$c \leq c^{PD}(\pi) \equiv \frac{L[\pi B + (1-\pi)b]}{L + 1 - \pi}.$$

Since

$$c^{PD}(\pi) - c^*(\pi) = \frac{(L-1)(1-\pi)[\pi B + (1-\pi)b]}{(L+1-\pi)(2-\pi)},$$

it follows that $c^{PD}(\pi) > c^*(\pi)$. Moreover, we can show that $\frac{\partial}{\partial \pi} c^{PD}(\pi) = L \frac{L(B-b)+B}{(L+1-p)^2} > 0$ and that $c^{PD} = b$ when $\pi = \frac{b}{L(B-b)+b}$.

3.9.2 PROOF OF PROPOSITION 3.5.1

Given a profile of tax rates $\boldsymbol{\tau} = \{\tau_1, ..., \tau_L\}$, a household resident in region ℓ decides where to invest its capital endowment by solving the following problem:

$$\max_{h_\ell, \{f_{\ell m}\}_{m \neq \ell}} h_\ell (1 - \tau_\ell) + \sum_{m \neq \ell} f_{\ell m} (1 - \tau_m) - \frac{1}{2} \left(\sum_{m \neq \ell} f_{\ell m} \right)^2$$

subject to its portfolio constraint

$$h_\ell + \sum_{m \neq \ell} f_{\ell m} = \kappa$$

and $(L-1)$ non-negativity constraints

$$f_{\ell m} \geq 0 \qquad \forall m \neq \ell,$$

where h_ℓ is capital invested in region ℓ, and $f_{\ell m}$ is capital invested in region $m \neq \ell$. Denote by $\lambda_{\ell m}$ the multipliers associated with the non-negativity constraints. Using the portfolio constraint to replace h_ℓ in the maximand of the household's problem, we obtain the first-order conditions for $f_{\ell m}$ and the complementary slackness conditions

$$\begin{cases} \tau_\ell - \tau_m + \lambda_{\ell m} = \sum\limits_{m \neq \ell} f_{\ell m} \\ \\ \lambda_{\ell m} f_{\ell m} = 0 \quad \lambda_{\ell m} \geq 0 \quad \forall m \neq \ell. \end{cases}$$

The proof of the proposition uses the following two lemmas.

Lemma 3.9.1. *Assume that there are two regions $m, n \neq \ell$, with $\tau_m > \tau_n$. Then $f_{\ell m} = 0$.*

Proof. Subtracting m's first-order condition from n's first-order condition, we obtain

$$\tau_m - \tau_n + \lambda_{\ell n} = \lambda_{\ell m}.$$

As $\tau_m > \tau_n$ and $\lambda_{\ell n} \geq 0$, $\lambda_{\ell m} > 0$. Hence, by the corresponding complementary slackness condition, $f_{\ell m} = 0$. □

Let $\widetilde{M}$ be the set of regions $\widetilde{m} \neq \ell$ that have chosen the minimum tax rate $\widetilde{\tau}_\ell = \min\{\tau_m\}_{m \neq \ell}$. Then, as an immediate consequence of Lemma 1, for all regions $m \neq \ell, \widetilde{m}$, $f_{\ell m} = 0$. Hence, tax rates τ_m become redundant; from now on, pertinent comparisons should be done only between τ_ℓ and $\widetilde{\tau}_\ell$.

Lemma 3.9.2. *Assume that $\tau_\ell \geq \widetilde{\tau}_\ell$. If $\widetilde{m} \in \widetilde{M}$ then $\lambda_{\ell \widetilde{m}} = 0$.*

Proof. We consider two cases: (i) Card$\{\widetilde{M}\} = 1$, (ii) Card$\{\widetilde{M}\} \geq 2$.
(i) First, assume that Card$\{\widetilde{M}\} = 1$ and consider the first-order condition

$$\tau_\ell - \widetilde{\tau}_\ell + \lambda_{\ell \widetilde{m}} = f_{\ell \widetilde{m}}. \tag{3.5}$$

If $\tau_\ell > \widetilde{\tau}_\ell$, as $\lambda_{\ell \widetilde{m}} \geq 0$, then $f_{\ell \widetilde{m}} > 0$. Thus, by the complementary slackness condition, $\lambda_{\ell \widetilde{m}} = 0$. If $\tau_\ell = \widetilde{\tau}_\ell$, (3.5) becomes

$$\lambda_{\ell \widetilde{m}} = f_{\ell \widetilde{m}}.$$

If $\lambda_{\ell \widetilde{m}} > 0$, then $f_{\ell \widetilde{m}} > 0$, which implies that $\lambda_{\ell \widetilde{m}} = 0$, which is a contradiction. Hence $\lambda_{\ell \widetilde{m}} = 0$.

(ii) Now assume that Card$\{\widetilde{M}\} \geq 2$. First-order conditions that characterize flows $f_{\ell\tilde{m}}$ are

$$\tau_\ell - \widetilde{\tau}_\ell + \lambda_{\ell\tilde{m}} = \sum_{\tilde{m}\in\widetilde{M}} f_{\ell\tilde{m}}.$$

In order to satisfy these first-order conditions, all multipliers $\lambda_{\ell\tilde{m}}$ should have the same value. If they were all strictly positive, then all outflows $f_{\ell\tilde{m}}$ should be equal to 0, implying that $\sum_{\tilde{m}\in\widetilde{M}} f_{\ell\tilde{m}} = 0$. But, as $\tau_\ell \geq \widetilde{\tau}_\ell$, all first-order conditions would yield a contradiction. Hence, all multipliers are zero. □

Therefore, from any first-order condition that characterizes a flow to a region $\tilde{m} \in \widetilde{M}$, we obtain

$$\sum_{\tilde{m}\in\widetilde{M}} f_{\ell\tilde{m}} = \tau_\ell - \widetilde{\tau}_\ell. \tag{3.6}$$

3.9.3 PROOF OF PROPOSITION 3.5.2

To obtain the equilibrium tax rates, first we derive region ℓ's reaction function.[33]

SCENARIO 1: ALL REGIONS HAVE DECIDED TO REFINANCE THEIR INCOMPLETE PROJECT

Denote by $\boldsymbol{\tau_m}$ the profile of tax rates chosen by regions $m \neq \ell$. For any profile $\boldsymbol{\tau_m}$, we need to consider three cases.

1. If the regional government of ℓ plans to set its tax rate strictly above $\widetilde{\tau}_\ell$, there will be capital outflows to regions $\tilde{m} \in \widetilde{M}$. Hence, the regional welfare would be

$$W_\ell^{FD} = (\kappa - \sum_{\tilde{m}\in\widetilde{M}} f_{\ell\tilde{m}})\,(1-\tau_\ell) + \sum_{\tilde{m}\in\widetilde{M}} f_{\ell\tilde{m}}\,(1-\widetilde{\tau}_\ell) - \frac{1}{2}(\sum_{\tilde{m}\in\widetilde{M}} f_{\ell\tilde{m}})^2 + b.$$

   By the Envelope Theorem, $\partial W_\ell^{FD}/\partial\tau_\ell = -(\kappa - \sum_{\tilde{m}\in\widetilde{M}} f_{\ell\tilde{m}}) < 0$. So the regional government of ℓ should set the lowest tax rate that satisfies its budget constraint

$$\theta\tau_\ell(\kappa - \sum_{\tilde{m}\in\widetilde{M}} f_{\ell\tilde{m}}) = c. \tag{3.7}$$

   Using (3.6), the smallest root of (3.7) is given by

$$\overline{\tau}_\ell = \frac{1}{2}\left[\kappa + \widetilde{\tau}_\ell - \sqrt{(\kappa+\widetilde{\tau}_\ell)^2 - \frac{4c}{\theta}}\right]. \tag{3.8}$$

[33] Due to specific features of this model, we cannot apply the methodology in Wildasin (1988) to derive the equilibrium tax rates.

Throughout the paper, we assume that κ is so large so that this square root always exists (see footnote 34).

2. If the regional government of ℓ plans to set its tax rate strictly below $\widetilde{\tau}_\ell$, there will be no capital outflows to regions $m \neq \ell$. Thus, regional welfare would be

$$W_\ell^{FD} = \kappa\,(1-\tau_\ell) + b.$$

Again, by the Envelope Theorem, $\partial W_\ell^{FD}/\partial\tau_\ell = -\kappa < 0$. So, regional government of ℓ should choose the lowest tax rate that satisfies its budget constraint

$$\theta\tau_\ell(\kappa + \sum_{m\neq\ell} f_{m\ell}) = c, \tag{3.9}$$

where

$$\sum_{m\neq\ell} f_{m\ell} = \sum_{m\neq\ell}\tau_m - (L-1)\tau_\ell \tag{3.10}$$

represents all capital inflows that leave regions $m \neq \ell$, and enter region ℓ. Rearranging terms, the smallest root of (3.9) is given by

$$\underline{\tau}_\ell = \frac{1}{2}\left[\frac{1}{L-1}\left(\kappa + \sum_{m\neq\ell}\tau_m\right) - \sqrt{\left(\frac{\kappa + \sum\limits_{m\neq\ell}\tau_m}{L-1}\right)^2 - \frac{4c}{\theta(L-1)}}\right]. \tag{3.11}$$

3. If the regional government of ℓ plans to replicate $\widetilde{\tau}_\ell$, there will be capital outflows from regions $m \notin \widetilde{M}$ to regions $\widetilde{m} \in \widetilde{M}$. Since $\ell \in \widetilde{M}$, the regional welfare of ℓ would be

$$W_\ell^{FD} = \kappa\,(1-\widetilde{\tau}_\ell) + \mathbb{1}_{\{ETC\geq c\}} b,$$

where

$$ETC = \theta\widetilde{\tau}_\ell(\kappa + \sigma_\ell \sum_{m\notin\widetilde{M}}\sum_{\widetilde{m}\in\widetilde{M}} f_{m\widetilde{m}}),$$

is the effective tax collection, and σ_ℓ, $0 < \sigma_\ell < 1$, represents the fraction of all capital outflows that leave regions $m \notin \widetilde{M}$ and move to region ℓ.

Now, in order to completely characterize the reaction function $\tau_\ell\,(\boldsymbol{\tau_m})$, we need to identify profiles of tax rates $\boldsymbol{\tau_m}$ for which $\underline{\tau}_\ell$, $\overline{\tau}_\ell$ and $\widetilde{\tau}_\ell$ are region ℓ's best responses.

First suppose that, facing a profile of tax rates $\boldsymbol{\tau_m}$, region ℓ wishes to set $\underline{\tau}_\ell$. It is straightforward to show that, for any change in one particular tax rate τ_m, $m \neq \ell$ that does not modify $\widetilde{\tau}_\ell$,

$$\frac{\partial\underline{\tau}_\ell}{\partial\tau_m} < 0.$$

Therefore, given $\widetilde{\tau}_\ell$, the highest tax rate $\underline{\tau}_\ell < \widetilde{\tau}_\ell$ is set when all regions $m \neq \ell$ have chosen $\tau_m = \widetilde{\tau}_\ell$. Denote by $\underline{\tau}_\ell(\widetilde{\tau}_\ell)$ this tax rate, which is given by

$$\underline{\tau}_\ell(\widetilde{\tau}_\ell) = \frac{1}{2}\left[\frac{\kappa}{L-1} + \widetilde{\tau}_\ell - \sqrt{\left(\frac{\kappa}{L-1} + \widetilde{\tau}_\ell\right)^2 - \frac{4c}{\theta(L-1)}}\right].$$

We can show that

$$\frac{\partial \underline{\tau}_\ell(\widetilde{\tau}_\ell)}{\partial \widetilde{\tau}_\ell} < 0$$

and

$$\lim_{\widetilde{\tau}_\ell \downarrow c/\theta\kappa} \underline{\tau}_\ell(\widetilde{\tau}_\ell) = c/\theta\kappa.$$

If $\widetilde{\tau}_\ell \leq c/\theta\kappa$, $\underline{\tau}_\ell(\widetilde{\tau}_\ell) \geq \widetilde{\tau}_\ell$, contradicting the definition of $\underline{\tau}_\ell(\widetilde{\tau}_\ell)$ as the highest tax rate that region ℓ can set strictly below $\widetilde{\tau}_\ell$. Therefore, if $\widetilde{\tau}_\ell < c/\theta\kappa$, the government in region ℓ cannot set a tax rate strictly less than $\widetilde{\tau}_\ell$ while, at the same time, satisfying its budget constraint. A similar argument applies if the regional government wishes to set $\tau_\ell = \widetilde{\tau}_\ell < c/\theta\kappa$. If this were the case, the effective tax collection would be

$$ETC = \theta\widetilde{\tau}_\ell\left(\kappa + \sigma_\ell \sum_{m \notin \widetilde{M}} \sum_{\widetilde{m} \in \widetilde{M}} f_{m\widetilde{m}}\right).$$

Consider the most favorable case for region ℓ: if $\sigma_\ell = 1$ and $(L-2)$ regions have set a tax rate equal to one, region ℓ receives the largest capital flow. Under this circumstance,

$$ETC < \frac{c}{\kappa}\left(\kappa + (L-2)(1-\widetilde{\tau}_\ell)\right) \simeq c$$

because $\tau_\ell < c/\theta\kappa$ and κ is assumed to be very large. Hence, when $\widetilde{\tau}_\ell < c/\theta\kappa$, the regional government of ℓ can only set a tax rate higher than $\widetilde{\tau}_\ell$ to cover the refinancing cost.

Now consider that, facing a profile of tax rates $\boldsymbol{\tau_m}$, region ℓ wishes to set $\overline{\tau}_\ell > \widetilde{\tau}_\ell$. It is straightforward to show that

$$\lim_{\widetilde{\tau}_\ell \to 0} \overline{\tau}_\ell = \frac{1}{2}\left[\kappa - \sqrt{\kappa^2 - \frac{4c}{\theta}}\right] \equiv \tau_\ell(0) > 0,$$

$$\frac{\partial \overline{\tau}_\ell}{\partial \widetilde{\tau}_\ell} < 0\,, \quad \frac{\partial^2 \overline{\tau}_\ell}{\partial \widetilde{\tau}_\ell{}^2} > 0$$

and

$$\lim_{\widetilde{\tau}_\ell \uparrow c/\theta\kappa} \overline{\tau}_\ell = c/\theta\kappa.$$

If $\widetilde{\tau}_\ell \geq c/\theta\kappa$, $\overline{\tau}_\ell(\widetilde{\tau}_\ell) \leq \widetilde{\tau}_\ell$, which again contradicts the definition of $\overline{\tau}_\ell$ as the lowest tax rate that region ℓ can set strictly above $\widetilde{\tau}_\ell$. Hence, if $\widetilde{\tau}_\ell > c/\theta\kappa$, the government in region ℓ cannot set a tax higher than $\widetilde{\tau}_\ell$ while, at the same time, satisfying its budget constraint. If the regional government sets $\tau_\ell = \widetilde{\tau}_\ell > c/\theta\kappa$, there will be no outflows $f_{\ell m}$. Thus, region ℓ's welfare will

amount to $W_\ell^{FD} = \kappa - c$. But this region will receive a capital inflow $\sigma_\ell \sum_{m \notin \widetilde{M}} \sum_{\widetilde{m} \in \widetilde{M}} f_{m\widetilde{m}}$, and so its tax collection will be

$$\theta \frac{c}{\theta\kappa}(\kappa + \sigma_\ell \sum_{m \notin \widetilde{M}} \sum_{\widetilde{m} \in \widetilde{M}} f_{m\widetilde{m}}) > c.$$

Therefore, there will be room for a decrease in τ_ℓ. Indeed, region ℓ can set its tax rate $\underline{\tau}_\ell < c/\theta\kappa$, satisfying its budget constraint and increasing its welfare (with respect to the choice of $\tau_\ell = \widetilde{\tau}_\ell > c/\theta\kappa$).

Two more cases remain to be analyzed. First, consider a profile of tax rates $\boldsymbol{\tau_m}$ such that $\widetilde{\tau}_\ell = c/\theta\kappa$ but when at least one region $n \neq \ell$ has set $\tau_n > c/\theta\kappa$. By a similar argument than before, region ℓ can set its tax rate $\underline{\tau}_\ell < c/\theta\kappa$, satisfying its budget constraint and increasing its welfare (with respect to the choice of $\tau_\ell = \widetilde{\tau}_\ell = c/\theta\kappa$.

Finally, when all regions $m \neq \ell$ have set the same tax rate $\tau_m = c/\theta\kappa$, we have already shown that region ℓ cannot tax strictly above or below $c/\theta\kappa$. Hence, the optimal response is to set $\tau_\ell = c/\theta\kappa$.

Region ℓ's reaction function is thus characterized as follows:

$$\tau_\ell\left(\boldsymbol{\tau_m}\right) = \begin{cases} \frac{1}{2}\left[\kappa + \widetilde{\tau}_\ell - \sqrt{\left(\kappa + \widetilde{\tau}_\ell\right)^2 - \frac{4c}{\theta}}\right] & \text{if } \tau_m < c/\theta\kappa \ \forall m \neq \ell \\ c/\theta\kappa & \text{if } \tau_m = c/\theta\kappa \ \forall m \neq \ell \\ \frac{1}{2}\left[\frac{\kappa + \sum_{m \neq \ell} \tau_m}{L-1} - \sqrt{\left(\frac{\kappa + \sum_{m \neq \ell} \tau_m}{L-1}\right)^2 - \frac{4c}{\theta(L-1)}}\right] & \text{if } \tau_m > c/\theta\kappa \ \forall m \neq \ell \text{ or} \\ & \tau_m = c/\theta\kappa \text{and } \exists n \neq \ell : \tau_n > c/\theta\kappa. \end{cases} \tag{3.12}$$

Region ℓ's reaction function is non-continuous. When $\widetilde{\tau}_\ell$ converges to $c/\theta\kappa$ from below, $\tau_\ell\left(\boldsymbol{\tau_m}\right)$ converges to this limit from above. But when the distribution of taxes is such that $\widetilde{\tau}_\ell = c/\theta\kappa$ but at least one region $n \neq \ell$ has set $\tau_n > c/\theta\kappa$, the optimal response is to set $\tau_\ell\left(\boldsymbol{\tau_m}\right) < c/\theta\kappa$.

Clearly, $\widehat{\tau}_1 = ... = \widehat{\tau}_L = c/\theta\kappa$ is a Nash equilibrium of this subgame because it is a fixed point of the best response correspondence. To prove uniqueness, we proceed in two steps. First, by simple inspection of (3.12), it is immediate that an asymmetric choice of taxes cannot be a Nash equilibrium. Second, there cannot be another symmetric equilibrium. Assume the contrary: let $\widehat{\tau}'_\ell = \widehat{\tau}'_m$ and, without any loss of generality,

$$\widehat{\tau}'_\ell = \widehat{\tau}'_m = \frac{c}{\theta\kappa} + \varepsilon, \tag{3.13}$$

with $\varepsilon \neq 0$ be another symmetric equilibrium. By substituting (3.13) into (3.12), we obtain a contradiction.

SCENARIO 2: AT LEAST ONE REGION HAS DECIDED NOT TO REFINANCE

In this case, to refinance its incomplete project the regional government of ℓ has to set the tax rate $\widehat{\tau}_\ell(0)$. This tax rate is obtained replacing $\widetilde{\tau}_\ell$ by 0 in the definition of $\overline{\tau}_\ell$.[34]

3.9.4 PROOF OF PROPOSITION 3.5.3

First, we prove the existence of c_1 and c_2. When there is at least one region that does not need refinancing, the total cost from completing a project in any region $T(c,\theta)$ is a strictly increasing and convex function of c, that satisfies

$$\lim_{c\to 0} T(c,\theta) = 0 \text{ and } \lim_{c\to b} T(c,\theta) = \frac{b}{\theta} + \frac{\left[\lim_{c\to b}\widehat{\tau}_\ell(0)\right]^2}{2} > b.$$

Hence, by Bolzano's Theorem, there exists a threshold $0 < c_1 < b$ such that, when $c \leq c_1$, $b - T(c,\theta) \geq 0$. Also, as $c/\theta > c$, there exists a threshold c_2 such that $c_2/\theta = b$. Moreover, as $c/\theta < T(c,\theta)$, $c_1 < c_2$.

Now consider the first case, when all regions face an incomplete project. When $0 \leq c \leq c_1$,

$$W_\ell^{FD}(r_\ell = R, \mathbf{r_m}, \mathbf{i}) = \kappa + b - c/\theta - c \geq W_\ell^{FD}(r_\ell = NR, \mathbf{r_m}, \mathbf{i}) = \kappa - c.$$

Here refinancing is a dominant strategy.

When $c_1 < c \leq c_2$,

$$W_\ell^{FD}(r_\ell = R, r_m = R, \mathbf{i}) = \kappa + b - c/\theta - c \geq W_\ell^{FD}(r_\ell = NR, r_m = R, \mathbf{i}) = \kappa - c \qquad \forall m \neq \ell$$

but

$$W_\ell^{FD}(r_\ell = R, r_m = NR, \mathbf{i}) = \kappa + b - T(c,\theta) - c \leq W_\ell^{FD}(r_\ell = NR, r_m = NR, \mathbf{i}) = \kappa - c \qquad \forall m \neq \ell.$$

These payoffs define a coordination (sub)game between regions, with two Nash equilibria. In the first equilibrium all regions refinance, while in the second one no region refinances.

In fact, we prove that the equilibrium where all regions refinance is both the strong (Aumann (1959) and the coalition-proof Nash equilibrium (Bernheim et al. (1987), as follows.

1. As both equilibria are Nash equilibria, no region can do better by unilaterally changing its equilibrium strategy. Then consider $\mathcal{L}-$regions coalitions, with $1 < \mathcal{L} < L$. If the other regions refinance (not refinance), the $\mathcal{L}$ regions do not want to deviate because $b - c/\theta \geq 0$ ($b - T(c,\theta) < 0$). Finally, consider the L-regions coalition. If all regions refinance, they do not want to deviate since $b - c/\theta \geq 0$. But, when no region refinances, they all wish to

[34] A sufficient condition for the existence of the square roots in (3.8) and (3.11), and to ensure that $\widehat{\tau}_\ell < 1$ is $\kappa \geq \max\{2\sqrt{b/\theta}, b/\theta, (L-1)4c/\theta\}$.

deviate because the first Nash equilibrium is Pareto optimal. Hence, the unique equilibrium that is strong is the first one.

2. Again, as both equilibria are Nash equilibria, no region can do better by unilaterally changing its equilibrium strategy. Then consider $\mathcal{L}-$regions coalitions, with $1 < \mathcal{L} < L$. If the other regions refinance, the coalition of $\mathcal{L}$ regions can jointly decide not to refinance. Although this deviation is self-enforcing, it is not worthy because $b - c/\theta \geq 0$. But if the other regions do not refinance, the unique available deviation to the coalition is to refinance. However this deviation is not self-enforcing because $b - T(c, \theta) < 0$. Finally, consider the L-regions coalition. If all regions refinance, they can jointly decide to not refinance. Although this deviation is self-enforcing, it is not worthy because $b - c/\theta \geq 0$. But when no region refinances, they all wish to deviate because the first Nash equilibrium is Pareto optimal. Hence, the unique coalition-proof Nash equilibrium is the first one.

Therefore, when $c_1 < c \leq c_2$, we choose the equilibrium where all regions refinance as the Nash equilibrium of this subgame.

Finally, when $c_2 < c \leq b$,

$$W_\ell^{FD}(r_\ell = R, \mathbf{r_m}, \mathbf{i}) = \kappa + b - c/\theta - c \leq W_\ell^{FD}(r_\ell = NR, \mathbf{r_m}, \mathbf{i}) = \kappa - c.$$

So not refinancing is a dominant strategy.

The proof of the second part of the proposition is immediate, and thus omitted.

3.9.5 PROOF OF PROPOSITION 3.5.4

First, we proceed to evaluate net expected regional welfares under different parameter conditions. If $c \leq c_1$, region ℓ's net expected welfare is:

$$\begin{aligned}
\mathbb{E}W_\ell^{FD}(i_\ell = I, i_m = I) &= \kappa - c + \pi B + (1-\pi)^L[b - \tfrac{c}{\theta}] \\
&\quad -(1-\pi)(1-(1-\pi)^{L-1})[b - T(c,\theta)] && \text{if } i_m = I \ \forall m \neq \ell \\
\mathbb{E}W_\ell^{FD}(i_\ell = I, i_m = NI) &= \kappa - c + \pi B + (1-\pi)[b - T(c,\theta)] && \text{if } \exists m \neq \ell : i_m = NI_\ell \\
\mathbb{E}W_\ell^{FD}(i_\ell = NI, i_m) &= \kappa && \forall i_m \in \{I, NI\}
\end{aligned}$$

If $c_1 < c \leq c_2$, region ℓ's net expected welfare is:

$$\begin{aligned}
\mathbb{E}W_\ell^{FD}(i_\ell = I, i_m = NI) &= \kappa - c + \pi B + (1-\pi)^L[b - \tfrac{c}{\theta}] && \text{if } i_m = I \ \forall m \neq \ell \\
\mathbb{E}W_\ell^{FD}(i_\ell = I, i_m = NI) &= \kappa - c + \pi B && \text{if } \exists m \neq \ell : i_m = NI_\ell \\
\mathbb{E}W_\ell^{FD}(i_\ell = NI, i_m) &= \kappa && \forall i_m \in \{I, NI\}
\end{aligned}$$

When $c_2 < c \leq b$, region ℓ's net expected welfare is:

$$\mathbb{E}W_\ell^{FD}(i_\ell = I, i_m) = \kappa - c + \pi B \quad \forall i_m \in \{I, NI\}$$
$$\mathbb{E}W_\ell^{FD}(i_\ell = NI, i_m) = \kappa \qquad\qquad \forall i_m \in \{I, NI\}$$

In order to characterize the Nash equilibria, we need to evaluate these different levels of regional welfare. We proceed as follows. First, define $\Delta^*(c) \equiv (1-\pi)[B-(b-c)]$. $\Delta^*(c)$ is an increasing, linear function of c that satisfies

$$\lim_{c\to 0} \Delta^*(c) = \Delta^*(0) \equiv (1-\pi)[B-b] \quad \text{and} \quad \lim_{c\to b} \Delta^*(c) = \Delta^*(b) \equiv (1-\pi)B.$$

Then, when $c \leq c_1$, we can show that:

1. $\mathbb{E}W_\ell^{FD}(i_\ell = I, i_m = I) \geq \mathbb{E}W_\ell^{FD}(i_\ell = NI, i_m) \Leftrightarrow \Delta_R^{FD}(c) \leq B - c$, where

$$\Delta_R^{FD}(c) \equiv (1-\pi)\left[B - (1-\pi)^{L-1}\left(b - \frac{c}{\theta}\right) - (1-(1-\pi)^{L-1})(b - T(c,\theta))\right]$$

   is a continuous, increasing, convex function of c that satisfies

$$\lim_{c\to 0} \Delta_R^{FD}(c) \equiv \ \Delta_R^{FD}(0) = (1-\pi)[B-b] = \Delta^*(0) \quad , \quad \Delta_R^{FD}(c) > \Delta^*(c)$$

   and

$$\lim_{c\to c_1} \Delta_R^{FD}(c) = \Delta_R^{FD}(c_1) \equiv (1-\pi)\left[B - (1-\pi)^{L-1}\left(b - \frac{c_1}{\theta}\right)\right] < \Delta^*(b).$$

2. $\mathbb{E}W_\ell^{FD}(i_\ell = I, i_m = NI) \geq \mathbb{E}W_\ell^{FD}(i_\ell = NI, i_m) \Leftrightarrow \Delta'(c) \leq B - c$, where

$$\Delta'(c) \equiv (1-\pi)[B - (b - T(c,\theta))]$$

   is another continuous, increasing, convex function of c that satisfies

$$\lim_{c\to 0} \Delta'(c) = (1-\pi)[B-b] = \Delta^*(0) \quad , \quad \Delta'(c) > \Delta_R^{FD}(c) > \Delta^*(c)$$

   and

$$\lim_{c\to c_1} \Delta'(c) = \Delta'(c_1) \equiv (1-\pi)B = \Delta^*(b).$$

When $c_1 < c \leq c_2$, we can also show that:

1. $\mathbb{E}W_\ell^{FD}(i_\ell = I, i_m = I) \geq \mathbb{E}W_\ell^{FD}(i_\ell = NI, i_m) \Leftrightarrow \Delta_{NAR}^{FD}(c) \leq B - c$, where

$$\Delta_{NAR}^{FD}(c) \equiv (1-\pi)\left[B - (1-\pi)^{L-1}\left(b - \frac{c}{\theta}\right)\right]$$

is a continuous, increasing, linear function of c that satisfies

$$\lim_{c \to c_1} \Delta_{NAR}^{FD}(c) = \Delta_R^{FD}(c_1) \quad , \quad \Delta_{NAR}^{FD}(c) > \Delta^*(c)$$

and

$$\lim_{c \to c_2} \Delta_{NAR}^{FD}(c) = \Delta_{NAR}^{FD}(c_2) \equiv (1-\pi)B = \Delta^*(b).$$

2. $\mathbb{E}W_\ell^{FD}(i_\ell = I, i_m = NI) \geq \mathbb{E}W_\ell^{FD}(i_\ell = NI, i_m) \Leftrightarrow \Delta''(c) \leq B - c$, where

$$\Delta''(c) \equiv (1-\pi)B$$

satisfies

$$\Delta''(c) = \Delta_{NAR}^{FD}(c_2) = \Delta'(c_1) = \Delta^*(b).$$

When $c_2 < c \leq b$, we show that $\mathbb{E}W_\ell^{FD}(i_\ell = I, i_m) \geq \mathbb{E}W_\ell^{FD}(i_\ell = NI, i_m) \Leftrightarrow \Delta_{NR}^{FD}(c) \leq B - c$, where

$$\Delta_{NR}^{FD}(c) \equiv (1-\pi)B = \Delta''(c).$$

The functions $\Delta^*(c)$, $\Delta_R^{FD}(c)$, $\Delta'(c)$, $\Delta_{NAR}^{FD}(c)$ and $\Delta_{NR}^{FD}(c)$ are (weakly) increasing in c, while $B - c$ decreases with c. Therefore, at some point, these five functions intersect $B - c$. Next, we first characterize these intersections as cost thresholds and divide the cost range $[0, b]$ in sub-intervals, according to these thresholds. Then we find the Nash equilibria in each corresponding sub-interval.

1. $0 \leq \pi \leq \pi_1^{FD}$

   Let π_1^{FD} be implicitly defined by $\Delta_R^{FD}(c_1) = B - c_1$. As $\Delta'(c) \geq \Delta_R^{FD}(c) \geq \Delta^*(c)$, the intersection between $\Delta'(c)$ and $B - c$ defines a threshold $c'(\pi) \leq c^*(\pi)$, whereas the intersection between $\Delta_R^{FD}(c)$ and $B - c$ defines another threshold $c_R^{FD}(\pi)$ that satisfies $c'(\pi) \leq c_R^{FD}(\pi) \leq c^*(\pi)$. The Nash equilibria are the following:[35]

   (a) When $0 \leq c \leq c'(\pi)$, $\Delta_R^{FD}(c) \leq \Delta'(c) \leq B - c$, which implies that

   $$\mathbb{E}W_\ell^{FD}(i_\ell = I, i_m = I) \geq \mathbb{E}W_\ell^{FD}(i_\ell = NI, i_m)$$

   and

   $$\mathbb{E}W_\ell^{FD}(i_\ell = I, i_m = NI) \geq \mathbb{E}W_\ell^{FD}(i_\ell = NI, i_m).$$

   Hence all regions initiate their project and, if it remains incomplete at the end of $t = 2$, they refinance it in $t = 3$.

   (b) When $c'(\pi) < c \leq c_R^{FD}(\pi)$, $\Delta_R^{FD}(c) \leq B - c < \Delta'(c)$, which implies that

   $$\mathbb{E}W_\ell^{FD}(i_\ell = I, i_m = I) \geq \mathbb{E}W_\ell^{FD}(i_\ell = NI, i_m)$$

[35] We only present the complete proof for this case. The remaining cases are analogous.

but

$$\mathbb{E}W_\ell^{FD}(i_\ell = I, i_m = NI) \leq \mathbb{E}W_\ell^{FD}(i_\ell = NI, i_m).$$

Two Nash equilibria emerge: i) all regions initiate their project and, if it remains incomplete at the end of $t = 2$, they refinance it in $t = 3$, or ii) no region invests. We choose the first equilibrium as it is strong and coalition-proof.

(c) When $c_R^{FD} < c$, $B - c < \Delta_R^{FD}(c) < \Delta'(c)$, which implies that

$$\mathbb{E}W_\ell^{FD}(i_\ell = I, i_m = I) \leq \mathbb{E}W_\ell^{FD}(i_\ell = NI, i_m)$$

and

$$\mathbb{E}W_\ell^{FD}(i_\ell = I, i_m = NI) \leq \mathbb{E}W_\ell^{FD}(i_\ell = NI, i_m).$$

Hence, no region initiates a project.

2. $\pi_1^{FD} < \pi \leq \pi_2^{FD}$

Let $\pi_2^{FD} \equiv c_1/B$ and $c_{NAR}^{FD}(\pi) \leq c^*(\pi)$ be defined by the intersection between $\Delta_{NAR}^{FD}(c)$ and $B - c$. The Nash equilibria are the following:

(a) When $0 \leq c \leq c_1$, all regions initiate their project and, if it remains incomplete at the end of $t = 2$, they refinance it in $t = 3$.

(b) When $c_1 < c \leq c_{NAR}^{FD}(\pi)$, all regions initiate their project and, if it remains incomplete at the end of $t = 2$, they refinance it in $t = 3$, provided all other regions do the same.

(c) When $c_{NAR}^{FD}(\pi) < c \leq b$, no region initiates a project.

3. $\pi_2^{FD} < \pi \leq \pi_3^{FD} \equiv b/B$

Let $c_{NR}^{FD}(\pi) \equiv \pi B \leq c^*(\pi)$ be defined by the intersection between $\Delta_{NR}^{FD}(c)$ and $B - c$. The Nash equilibria are the following:

(a) When $0 \leq c \leq c_1$, all regions initiate their project and, if it remains incomplete at the end of $t = 2$, they refinance it in $t = 3$.

(b) When $c_1 < c \leq c_2$, all regions initiate their project and, if it remains incomplete at the end of $t = 2$, they refinance it in $t = 3$, provided all other regions do the same.

(c) When $c_2 < c \leq c_{NR}^{FD}(\pi)$, all regions initiate their project but, if it remains incomplete at the end of $t = 2$, they do not refinance it.

(d) When $c_{NR}^{FD}(\pi) < c \leq b$, no region initiates a project.

4. $\pi_3^{FD} < \pi \leq 1$

The Nash equilibria are the following:

(a) When $0 \leq c \leq c_1$, all regions initiate their project and, if it remains incomplete at the end of $t = 2$, they refinance it in $t = 3$.

(b) When $c_1 < c \leq c_2$, all regions initiate their project and, if it remains incomplete at the end of $t = 2$, they refinance it in $t = 3$, provided all other regions do the same.

(c) When $c_2 < c \leq b$, all regions initiate their project but, if it remains incomplete at the end of $t = 2$, they do not refinance it.

3.9.6 PROOF OF PROPOSITION 3.6.1

The proof of this proposition uses the expected welfare expressions derived in the Online Appendix.

WELFARE ANALYSIS WHEN π ADOPTS EXTREME VALUES

We consider two cases: (i) $\pi = 0$, (ii) $b/B \leq \pi \leq 1$.

(i) $\underline{\pi = 0}$

When $\pi = 0$, $c^*(0) = b/2$, $c^{PD}(0) = \frac{Lb}{L+1}$ and $c_R^{FD}(0) = \frac{\theta b}{1+\theta}$. Hence, expected welfare under each regime is

$$\mathbb{E}W^{PD}(\pi) = \kappa + \frac{bL}{(1+L)^2} \quad \text{and} \quad \mathbb{E}W^{FD}(0,\theta) = \kappa + \frac{\theta b}{2(1+\theta)}.$$

When $\theta \geq \theta_0$, $\mathbb{E}W^{FD} \geq \mathbb{E}W^{PD}$. Otherwise, partial decentralization dominates.

Last, we analyze the value of $\left.\frac{\partial \mathbb{E}W^{FD}(\pi,\theta)}{\partial \pi}\right|_{\pi=0}$ as a function of the fiscal capacity θ. For any pair (b,θ), we can always find a real number β and write $\kappa = \beta\sqrt{\frac{4b}{1+\theta}}$. Computing the abovementioned derivative, we obtain

$$-\frac{\theta L[-12bB(1+\theta) + 8(L-1)\beta b^2(\beta^2-1)^{\frac{3}{2}} + b^2(9 - 12\beta^2 + 8\beta^4 + 12\theta + L(-8\beta^4 + 2\beta^2 - 3))]}{12(1+\theta)^2}.$$

The value of L that makes this expression equal to zero is

$$\tilde{L}(\theta) = 1 + \frac{3[4B(1+\theta) - 2b(1+2\theta)]}{b[-8\beta^4 + 12\beta^2 + 8\beta(\beta^2-1)^{\frac{3}{2}} - 3]} \geq 1.$$

The fraction's denominator converges fast to zero, i.e., for values of β below 1. But recall that in footnote 29 we have imposed κ to be sufficiently large; so β should be in fact a real number much larger than 1. Thus, $\tilde{L}(\theta)$ turns out to be a big number. Moreover, $\tilde{L}(\theta)$ increases with θ. Hence, if $L \leq \bar{L} \equiv \tilde{L}(0)$

$$\left.\frac{\partial \mathbb{E}W^{FD}(\pi,\theta)}{\partial \pi}\right|_{\pi=0} \geq 0, \forall \theta \in [0,1]$$

and it increases with θ. As we prove in the Online Appendix that $\mathbb{E}W^{FD}(\pi,\theta)$ is convex, we can conclude that $\mathbb{E}W^{FD}(\pi,\theta)$ is an increasing function of the administrative capacity π.

(ii) $\underline{\pi_2^{PD} = \pi_3^{FD} = b/B \leq \pi \leq 1}$

When $b/B \leq \pi < 1$, partial decentralization replicates the first best outcomes, whereas full decentralization's outcomes are distorted. Hence, partial decentralization dominates. But if we compute

$$\frac{\partial}{\partial \pi}\left[\mathbb{E}W^{FD}(\pi,\theta) - \mathbb{E}W^{PD}(\pi)\right]$$

and we take the limit when π converges to one, we obtain

$$\lim_{\pi \to 1} \frac{\partial}{\partial \pi}\left[\mathbb{E}W^{FD}(\pi,\theta) - \mathbb{E}W^{PD}(\pi)\right] = \frac{1}{b}\left(\int_0^b [b-c]dc - \int_0^{c_1} [b - T(c,\theta)]dc\right).$$

As $c_1 < b$ and $T(c,\theta) > c$, this limit is strictly positive. So, when π converges to 1, the full decentralization expected welfare converges, from below, to the first best level. When $\pi = 1$,

$$\mathbb{E}W^{PD}(1) = \mathbb{E}W^{FD}(\pi,\theta) = \kappa + B - \frac{b}{2}.$$

WELFARE COMPARISON WHEN $\theta \in [\theta^*, 1]$

First, assume that $\theta = 1$. When $\pi = 0$,

$$\mathbb{E}W^{FD}(0,1) = \frac{bL}{4},$$

and when $\pi = \pi_1^{PD}$,

$$\mathbb{E}W^{PD}(\pi_1^{PD}) = \frac{bL(2B-b)}{2[BL - b(L-1)]}.$$

If $L \geq \underline{L} \equiv 3 + \frac{B}{B-b}$, $\mathbb{E}W^{FD}(0,1) \geq \mathbb{E}W^{PD}(\pi_1^{PD})$. Hence, $\mathbb{E}W^{FD}(\pi,\theta)$ lies everywhere above $\mathbb{E}W^{PD}(\pi)$ when $\pi \in [0, \pi_1^{PD}]$ because the former is an increasing function of π. Therefore, as $\mathbb{E}W^{FD}(\pi,\theta)$ must converge to $\mathbb{E}W^{PD}(\pi)$ from below when π converges to one, $\mathbb{E}W^{FD}(\pi,\theta)$ has to cross $\mathbb{E}W^{PD}(\pi)$ in its linear part, from above. Denote by $\hat{\pi}(1)$ the administrative capacity level that corresponds to this intersection.[36] In fact, $\hat{\pi}(1)$ is unique. If this were not the case, $\mathbb{E}W^{FD}(\pi,\theta)$ would cross again $\mathbb{E}W^{PD}(\pi)$, from below. But, if such second intersection occurs, by convexity, $\mathbb{E}W^{FD}(\pi,\theta)$ could not converge to $\mathbb{E}W^{PD}(\pi)$ at $\pi = 1$.

When θ decreases, $\mathbb{E}W^{FD}(\pi,\theta)$ decreases as well, while $\mathbb{E}W^{PD}(\pi)$ remains constant. By continuity, we know that there exists θ^*, a value of the regional fiscal capacity such that $\mathbb{E}W^{FD}(\pi,\theta^*)$ lies above $\mathbb{E}W^{PD}(\pi)$ everywhere when $\pi \in [0, \pi_1^{PD}]$. Hence, when $\theta \in [\theta^*, 1]$, $\mathbb{E}W^{FD}(\pi,\theta)$ intersects $\mathbb{E}W^{PD}(\pi)$ in its linear part. Using the same geometrical argument as before, we know that both expected welfares cross only once, at $\hat{\pi}(\theta)$. Applying the Implicit Function Theorem, we show that

$$\frac{\partial \hat{\pi}(\theta)}{\partial \theta} = -\frac{\partial \mathbb{E}W^{FD}(\hat{\pi}(\theta),\theta)/\partial\theta}{\partial \mathbb{E}W^{FD}(\hat{\pi}(\theta),\theta)/\partial\pi - \partial \mathbb{E}W^{PD}(\hat{\pi}(\theta))/\partial\theta} > 0$$

[36] As partial decentralization dominates when $\pi \geq \pi_2^{PD}$, $\pi_1^{PD} \leq \hat{\pi}(1) < \pi_2^{PD}$.

because, at $\hat{\pi}(\theta)$, $\mathbb{E}W^{FD}(\pi,\theta)$ crosses $\mathbb{E}W^{PD}(\pi)$ from above. Thus, $\hat{\pi}(\theta)$ increases with θ.

When $\theta < \theta^*$, we cannot a priori ensure that $\mathbb{E}W^{FD}(\pi,\theta)$ crosses $\mathbb{E}W^{PD}(\pi)$ only once. The goal of the following paragraphs is to prove that this is indeed the case.

WELFARE COMPARISON WHEN $\theta \leq \theta_0$

When $\theta = \theta_0$,

$$\left.\frac{\partial \mathbb{E}W^{PD}(\pi)}{\partial \pi}\right|_{\pi=0} = \frac{L[4(L+1)B + (L^2-5L-2)b]}{2(L+1)^3} > 0$$

because $L^2 - 5L - 2 > -8.25$ and $L \geq \underline{L} \geq 3$. Also

$$\begin{aligned}
\left.\frac{\partial \mathbb{E}W^{FD}(\pi,\theta)}{\partial \pi}\right|_{\pi=0} &= \frac{1}{b}\int_0^{\frac{\theta_0 b}{1+\theta_0}} \left[B - b + L\frac{c}{\theta_0} + (1-L)T(c,\theta_0)\right] dc \\
&< \frac{1}{b}\left\{\int_0^{\frac{\theta_0 b}{1+\theta_0}} [B-b]dc + \int_0^{\frac{\theta_0 b}{1+\theta_0}} T(c,\theta_0)dc\right\} \\
&< \frac{1}{b}\int_0^{\frac{\theta_0 b}{1+\theta_0}} [B-b]dc + \frac{\theta_0 b}{2(1+\theta_0)} = \frac{(2B-b)L}{(1+L)^2},
\end{aligned}$$

because $T(c,\theta_0) \leq b$. Hence,

$$\left(\frac{\partial \mathbb{E}W^{PD}(\pi)}{\partial \pi} - \frac{\partial \mathbb{E}W^{FD}(\pi,\theta_0)}{\partial \pi}\right)\Bigg|_{\pi=0} > \frac{bL^2(L-3)}{2(L+1)^3} > 0.$$

Next, we prove that, when $\pi \in [0, \pi_1^{PD}]$, $\mathbb{E}W^{PD}(\pi) > \mathbb{E}W^{FD}(\pi,\theta_0)$. To do so, we first find an upper-bound for the expected welfare under full decentralization. If $\pi_1^{PD} \geq \pi_2^{FD}$,

$$\begin{aligned}
\mathbb{E}W^{FD}(\pi_1^{PD},\theta_0) = \kappa + \frac{1}{b}\Bigg\{&\int_0^{c_{NR}^{FD}(\pi)} \left[\pi_1^{PD}B - c\right] dc + \int_0^{c_2} \left[\left(1-\pi_1^{PD}\right)^L \left(b - \frac{c}{\theta_0}\right)\right] dc \\
&+ \int_0^{c_1} \left[\left(1-\pi_1^{PD}\right)\left(1 - \left(1-\pi_1^{PD}\right)^{L-1}\right)(b - T(c,\theta_0))\right] dc\Bigg\}
\end{aligned}$$

where $c_{NR}^{FD} = \pi_1^{PD}B$ and $c_2 = \theta_0 b$. So

$$\begin{aligned}
\mathbb{E}W^{FD}(\pi_1^{PD},\theta_0) &= \kappa + \frac{bB^2}{2[b+L(B-b)]^2} + \frac{bL(1-\frac{b}{b+L(B-b)})^L}{1+L^2} \\
&\quad + \frac{1}{b}\int_0^{c_1}\left[\left(1-\pi_1^{PD}\right)\left(1-\left(1-\pi_1^{PD}\right)^{L-1}\right)(b - T(c,\theta_0))\right] dc \\
&< \kappa + \frac{bB^2}{2[b+L(B-b)]^2} + \frac{bL(1-\frac{b}{b+L(B-b)})^L}{1+L^2}
\end{aligned}$$

$$+\frac{1}{b}\int_0^{c_1}\left[\left(1-\pi_1^{PD}\right)\left(1-\left(1-\pi_1^{PD}\right)^{L-1}\right)\left(b-\frac{c}{\theta_0}\right)\right]dc$$

$$=\kappa+\frac{bB^2}{2[b+L(B-b)]^2}+\frac{bL\left(1-\frac{b}{b+L(B-b)}\right)}{1+L^2}\equiv A.$$

If $\pi_1^{PD}<\pi_2^{FD}$,

$$\mathbb{E}W^{FD}(\pi_1^{PD},\theta_0)=\kappa+\frac{1}{b}\left\{\int_0^{c_{NAR}^{FD}(\pi)}\left[\pi_1^{PD}B-c+\left(1-\pi_1^{PD}\right)^L\left(b-\frac{c}{\theta_0}\right)\right]dc\right.$$

$$\left.+\int_0^{c_1}\left[\left(1-\pi_1^{PD}\right)\left(1-\left(1-\pi_1^{PD}\right)^{L-1}\right)(b-T(c,\theta_0))\right]dc\right\}$$

$$<\kappa+\frac{1}{b}\left\{\int_0^{c_2}\left[\pi_1^{PD}B-c+\left(1-\pi_1^{PD}\right)^L\left(b-\frac{c}{\theta_0}\right)\right]dc\right.$$

$$\left.\int_0^{c_1}\left[\left(1-\pi_1^{PD}\right)\left(1-\left(1-\pi_1^{PD}\right)^{L-1}\right)(b-T(c,\theta_0))\right]dc\right\}<A.$$

Hence A is one upper bound for $\mathbb{E}W^{FD}(\pi_1^{PD},\theta_0)$. Now, define

$$\Gamma_{\theta_0}(\pi_1^{PD}) \equiv \mathbb{E}W^{PD}(0)+\left.\frac{\partial\mathbb{E}W^{PD}(\pi)}{\partial\pi}\right|_{\pi=0}.\pi_1^{PD} =\kappa-\frac{bL[bL(L+5)-2B(L+1)(L+2)]}{2(1+L)^3[BL+b(L-1)]}.$$

Finally, we evaluate

$$\Gamma_{\theta_0}(\pi_1^{PD})-A=H(b,B,L)\equiv \frac{1}{2}b\left[\frac{-B^2}{[b+L(B-b)]^2}-\frac{L[bL(L+5)-2B(L+1)(L+2)]}{(1+L)^3[BL+b(L-1)]}-\frac{2L\left(1-\frac{b}{b+L(B-b)}\right)}{1+L^2}\right].$$

As $L\geq\underline{L}$, we can write $L=2+n+\frac{B}{B-b}$, for $n\geq 1$. The solution to the equation $H(b,B,L)=0$ in B can be expressed in the form $B=\alpha(n).b$. Although it is not possible to obtain a close form expression for $\alpha(n)$, the following figure depicts the curve $\alpha(n)$, when $n\in\{1,...,100\}$.

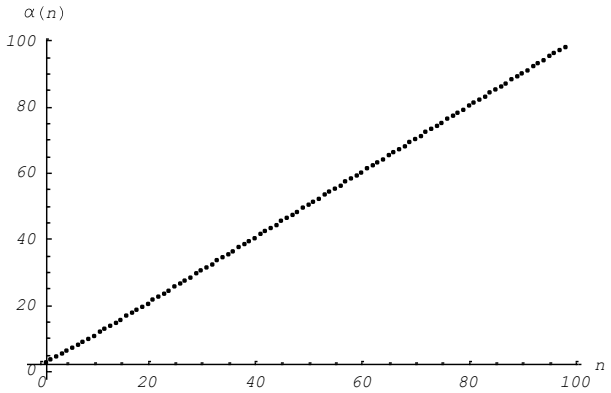


Figure 3.9: The Function $\alpha(n)$.

As $\alpha(1) = 2.6817$ and $B/2 < b$, $H(b, B, L) > 0$ for any $n \in \{1, ..., 100\}$.[37] Therefore, we conclude that $\mathbb{E}W^{FD}(\pi_1^{PD}, \theta_0) < \Gamma_0(\pi_1^{PD})$.

These results enable us to assert that, when $\pi \in [0, \pi_1^{PD}]$, $\mathbb{E}W^{FD}(\pi, \theta_0)$ lies below $\Gamma_{\theta_0}(\pi)$, a straight line that takes the values $\mathbb{E}W^{PD}(0)$ and $\Gamma_{\theta_0}(\pi_1^{PD})$ when $\pi = 0$ and π_1^{PD}, respectively. In fact, $\Gamma_{\theta_0}(\pi)$ is the tangent line to $\mathbb{E}W^{PD}(\pi,)$ at $\pi = 0$. Hence, as $\mathbb{E}W^{PD}(\pi)$ is convex, $\mathbb{E}W^{FD}(\pi, \theta_0) < \mathbb{E}W^{PD}(\pi)$ everywhere on $\pi \in [0, \pi_1^{PD}]$. We depict this result in Figure 3.10.

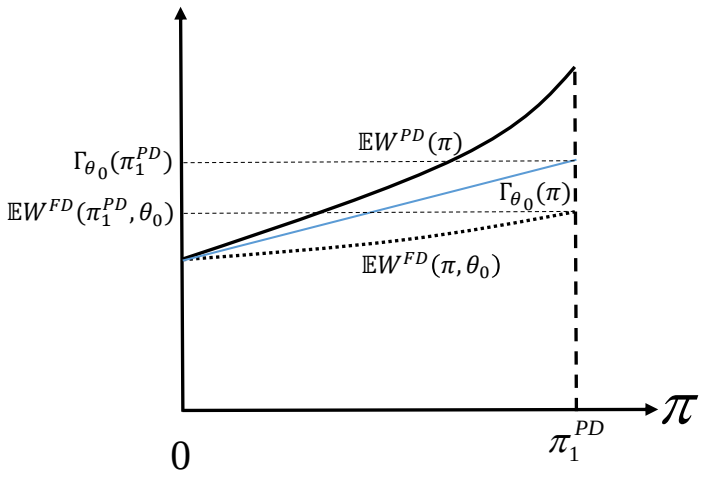


Figure 3.10: Expected Welfare when $\theta = \theta_0$ and $\pi \in [0, \pi_1^{PD}]$.

Moreover, when $\pi \in (\pi_1^{PD}, 1]$, $\mathbb{E}W^{FD}(\pi, \theta_0)$ cannot cross $\mathbb{E}W^{PD}(\pi)$. If this were the case, the former curve would intersect the latter from below, in its linear part. But then, by convexity, it could not converge towards $\mathbb{E}W^{PD}(\pi)$ from below at $\pi = 1$.

Hence, when $\theta = \theta_0$, $\mathbb{E}W^{FD}(\pi, \theta_0) < \mathbb{E}W^{PD}(\pi)$, except at $\pi = 0$ when they coincide. Therefore, as $\mathbb{E}W^{FD}(\pi, \theta)$ increases with θ, $\mathbb{E}W^{FD}(\pi, \theta) < \mathbb{E}W^{PD}(\pi)$ for all values of π whenever $\theta \leq \theta_0$.

I We will prove, using a series of geometrical arguments, that when $\theta \in [\theta_0, \theta^*]$, $\mathbb{E}W^{FD}(\pi, \theta)$ intersects $\mathbb{E}W^{PD}(\pi)$ exactly once.

Assume that we can find $\theta = \theta_0 + \epsilon$ such that

$$\mathbb{E}W^{FD}(0, \theta_0 + \epsilon) > \mathbb{E}W^{PD}(0) \text{ and } \mathbb{E}W^{FD}(\pi_1^{PD}, \theta_0 + \epsilon) < \Gamma_{\theta_0}(\pi_1^{PD}) < \mathbb{E}W^{PD}(\pi_1^{PD}).$$

Clearly, $\mathbb{E}W^{FD}(\pi, \theta_0 + \epsilon)$ intersects the line $\Gamma_{\theta_0}(\pi)$ once, from above. Hence, as $\mathbb{E}W^{FD}(\pi, \theta_0 + \epsilon)$ and $\mathbb{E}W^{PD}(\pi)$ are both convex in π, the former has to cross the latter only once, from above. This argument can be replicated until $\theta = \theta_1 > \theta_0$, which is implicitly defined by

$$\mathbb{E}W^{FD}(\pi_1^{PD}, \theta_1) = \Gamma_{\theta_0}(\pi_1^{PD}).$$

When $\theta = \theta_1$, there can be two possibilities. When π converges to π_1^{PD}, either (i) $\mathbb{E}W^{FD}(\pi, \theta_1)$ converges to $\Gamma_{\theta_0}(\pi_1^{PD})$ from below, in which case the previous intersection argument applies, or

[37] This result also holds for $n > 100$.

(ii) this convergence is from above. This means that

$$\left.\frac{\partial \mathbb{E}W^{FD}(\pi,\theta_1)}{\partial \pi}\right|_{\pi=\pi_1^{PD}} < \text{slope}\,\Gamma_{\theta_0}(\pi).$$

Hence, $\mathbb{E}W^{FD}(\pi,\theta_1)$ has to cross $\mathbb{E}W^{PD}(\pi)$ from above, only once. So, for any $\theta \in [\theta_0,\theta_1]$, $\mathbb{E}W^{FD}(\pi,\theta)$ intersects $\mathbb{E}W^{PD}(\pi)$ exactly once.

Now, let's define $\Gamma_{\theta_1}(\pi)$, a new line that has the same slope than $\Gamma_{\theta_0}(\pi)$ and is characterized by $\Gamma_{\theta_1}(0) = \mathbb{E}W^{FD}(0,\theta_1)$. By construction, $\Gamma_{\theta_1}(\pi_1^{PD}) > \mathbb{E}W^{FD}(\pi_1^{PD},\theta_1)$. Now assume that we can find $\theta = \theta_1 + \epsilon'$ such that $\mathbb{E}W^{FD}(0,\theta_1+\epsilon') > \Gamma_{\theta_1}(0)$ and $\mathbb{E}W^{FD}(\pi_1^{PD},\theta_1+\epsilon') < \Gamma_{\theta_1}(\pi_1^{PD})$. Clearly, $\mathbb{E}W^{FD}(\pi,\theta_1+\epsilon')$ crosses $\Gamma_{\theta_1}(\pi)$ only once, from above. Hence, as $\mathbb{E}W^{FD}(\pi,\theta_1+\epsilon)$ and $\mathbb{E}W^{PD}(\pi)$ are both convex in π, the former has to cross the latter only once, from above. This argument can be replicated until $\theta = \theta_2 > \theta_1$, which is implicitly defined by

$$\mathbb{E}W^{FD}(\pi_1^{PD},\theta_2) = \Gamma_{\theta_1}(\pi_1^{PD}).$$

Now construct an increasing sequence θ_n, $\forall n \geq 0$ defined by

$$\mathbb{E}W^{FD}(0,\theta_{n+1}) > \Gamma_{\theta_n}(0),$$

and

$$\mathbb{E}W^{FD}(\pi_1^{PD},\theta_{n+1}) = \Gamma_{\theta_n}(\pi_1^{PD}).$$

This sequence can behave in two different ways: either there exists $N \in \mathbb{N}$ such that, for all $n \geq N$, $\theta_n \geq \theta^*$; or $\theta_n < \theta^*$ for all $n \in \mathbb{N}$. In the first case, the previous geometric arguments apply, and thus we can assert that $\mathbb{E}W^{FD}(\pi,\theta)$ crosses $\mathbb{E}W^{PD}(\pi)$ only once, from above, when $\pi \in [0,\pi_1^{PD}]$.

Let's prove that the second case can be ruled out. Assume that $\theta_n < \theta^*$ for all $n \in \mathbb{N}$. As the sequence θ_n is increasing and bounded (by $\theta = 1$), it must converge. Denote this limit by $\overline{\theta}$. As the functions that define the sequence are continuous, they also converge towards $\mathbb{E}W^{FD}(\pi,\overline{\theta})$ and $\Gamma_{\overline{\theta}}(\pi)$. These limit functions have to satisfy

$$\mathbb{E}W^{FD}(0,\overline{\theta}) = \Gamma_{\overline{\theta}}(0),$$

and

$$\mathbb{E}W^{FD}(\pi_1^{PD},\overline{\theta}) = \Gamma_{\overline{\theta}}(\pi_1^{PD}).$$

Define $\overline{\overline{\theta}} = \overline{\theta} + \epsilon$. Clearly, as $\mathbb{E}W^{FD}(\pi,\theta)$ increases with θ, $\mathbb{E}W^{FD}(0,\overline{\overline{\theta}}) > \Gamma_{\overline{\theta}}(0)$. But if this were the case, we can construct $\Gamma_{\overline{\overline{\theta}}}(\pi) > \Gamma_{\overline{\theta}}(\pi)$, which is a contradiction. Hence, the sequence θ_n does not converge to a value $\overline{\theta} < \theta^*$.

We conclude that, when $\theta \in [\theta_0,\theta^*]$, $\mathbb{E}W^{FD}(\pi,\theta)$ intersects $\mathbb{E}W^{PD}(\pi)$ exactly once, from above, at $\hat{\pi}(\theta)$. Applying the Implicit Function Theorem, we can show that $\partial\hat{\pi}(\theta)/\partial\theta > 0$.

Bibliography

Abel, A. and M. Warshawsky (1988). Specification of the Joy of Giving: Insights from Altruism. *The Review of Economics and Statistics 70*, 145–149.

Abel, A. B. and L. J. Kotlikoff (1994). Intergenerational Altruism and the Effectiveness of Fiscal Policy–New Tests Based on Cohort Data. In T. Tachibanaki (Ed.), *Savings and Bequests*, Chapter 7, pp. 167 – 196.

Acemoğlu, D. and D. Autor (2011). Skills, Tasks and Technologies: Implications for Employment and Earnings. In D. Card and O. Ashenfelter (Eds.), *Handbook of Labor Economics*, Volume 4B, pp. 1043–1171. Elsevier Press.

Acemoğlu, D., C. Garcia-Jimeno, and J. Robinson (2014). State Capacity and Economic Development: A Network Approach. Working paper.

Akai, N. and M. Sato (2008). Too Big or Too Small? A Synthetic View of the Commitment Problem of Interregional Transfers. *Journal of Urban Economics 64*(3), 551–559.

Akerlof, G. A. (1978). The Economics of 'Tagging' As Applied to the Optimal Income Tax, Welfare Programs, and Manpower Planning. *American Economic Review 68*, 8–19.

Albanesi, S. (2011). Optimal Taxation of Entrepreneurial Capital with Private Information. Working paper.

Albanesi, S. and C. Sleet (2006). Dynamic Optimal Taxation with Private Information. *Review of Economic Studies 73*, 1–30.

Alder, S. (2012). In the Wrong Hands: Complementarities, Resource Allocation, and Aggregate TFP. Working paper.

Ales, L., M. Kurnaz, and C. Sleet (2014). Tasks, Talents and Taxes. Working paper.

Altonji, J. G. (1986). Intertemporal Substitution in Labor Supply: Evidence from Micro Data. *Journal of Political Economy 94*(3), 176–215.

Altonji, J. G., F. Hayashi, and L. J. Kotlikoff (1992). Is the Extended Family Altruistically Linked? Direct Tests Using Micro Data. *American Economic Review 82*, 1177–1198.

Altonji, J. G., F. Hayashi, and L. J. Kotlikoff (1997). Parental Altruism and Inter Vivos Transfers: Theory and Evidence. *Journal of Political Economy 105*, 1121–1166.

Arbetman-Rabinowitz, M., J. Kugler, M. Abdollahian, K. Kang, H. Nelson, and R. Tammen (2012). Political Performance. In J. Kugler and R. Tammen (Eds.), *The Performance of Nations*, pp. 7–48. Rowman and Littlefield Publishers.

Atkeson, A. and R. E. Lucas (1992). On Efficient Distribution with Private Information. *Review of Economic Studies 59*, 427–453.

Atkinson, A. B. and J. E. Stiglitz (1976). The Design of Tax Structure: Direct versus Indirect Taxation. *Journal of Public Economics 6*, 55–75.

Aumann, R. (1959). Acceptable Points in General Cooperative n-Person Games. In R. Luce and A. Tucker (Eds.), *Contributions to the Theory of Games*, pp. 287–324. Princeton University Press.

Axtell, R. L. (2001). Zipf distribution of US firm sizes. *Science 293*(5536), 1818–1820.

Baker, G. P., M. C. Jensen, and K. J. Murphy (1988). Compensation and Incentives: Practice vs. Theory. *The Journal of Finance 43*(3), 593–616.

Bakija, J., A. Cole, and B. Heim (2012). Jobs and Income Growth of Top Earners and the Causes of Changing Income Inequality: Evidence from U.S. Tax Return Data. Working paper.

Bandyopadhyay, S. and E. Green (2012). Pre-Colonial Political Centralization and Contemporary Development in Uganda. Working paper.

Bardhan, P. and D. Mookherjee (2006). Decentralization and Accountability in Infrastructure Delivery in Developing Countries. *The Economic Journal 116*(508), 101–127.

Bartelsman, E. J. and M. Doms (2000). Understanding productivity: Lessons from longitudinal microdata. *Journal of Economic Literature 38*(3), 569–594.

Becker, G. S. and N. Tomes (1986). Human Capital and the Rise and Fall of Families. *Journal of Labor Economics 4*, S1–S39.

Berman, E., J. Bound, and Z. Griliches (1994). Changes in the Demand for Skilled Labor within U.S. Manufacturing: Evidence from the Annual Survey of Manufactures. *The Quarterly Journal of Economics 109*(2), 367–397.

Bernheim, D., D. Peleg, and M. Whinston (1987). Coalition-Proof Nash Equilibria. I: Concepts. *Journal of Economic Theory 42*(1), 1–12.

Besfamille, M. and B. Lockwood (2008). Bailouts in Federations: Is a Hard Budget Constraint Always Best? *International Economic Review 49*(2), 577–593.

Besley, T. and T. Persson (2010). State Capacity, Conflict and Development. *Econometrica 78*(1), 1–34.

Bird, R. (1995). Decentralizing Infrastructure: For Good or for Ill? World Bank Discussion Papers.

Blundell, R., L. Pistaferri, and I. Saporta-Eksten (2012). Consumption Inequality and Family Labor Supply. Working paper.

Boadway, R. and A. Shah (2007). *Intergovernmental Fiscal Transfers Principles and Practice*. The World Bank.

Brown, J. R. (2000). Differential Mortality and the Value of Individual Account Retirement Annuities. In M. Feldstein and J. B. Liebman (Eds.), *The Distributional Aspects of Social Security and Social Security Reform*, pp. 401–446. University of Chicago Press.

Brueckner, J. (2008). Partial Fiscal Decentralization. *Regional Science and Urban Economics 39*(1), 23–32.

Bucovetsky, S. (1991). Asymmetric Tax Competition. *Journal of Urban Economics 30*(2), 167–181.

Cagetti, M. and M. D. Nardi (2006). Entrepreneurship, Frictions, and Wealth. *Journal of Political Economy 114*(5), 835–870.

Caplan, A., R. Cornes, and E. Silva (2000). Pure Public Goods and Income Redistribution in a Federation with Decentralized Leadership and Imperfect Labour Mobility. *Journal of Public Economics 77*(2), 265–284.

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.

Corbacho, A., V. Fretes Cibils, and E. Lora (2013). *More than Revenue: Taxation as a Development Tool*. Palgrave Macmillan.

Cremer, H., M. Marchand, and P. Pestieau (1997). Investment in Local Public Services: Nash Equilibrium and Social Optimum. *Journal of Public Economics 65*(1), 23–35.

Cremer, H. and P. Pestieau (2006). Wealth Transfer Taxation: A Survey of the Theoretical Literature. *Handbook of the Economics of Giving, Altruism and Reciprocity 2*, 1107–1134.

Crivelli, E. and K. Staal (2013). Size, Spillovers and Soft Budget Constraints. *International Tax and Public Finance 20*(1), 338–356.

De Nardi, M., E. French, and J. B. Jones (2006). Differential Mortality, Uncertain Medical Expenses, and the Saving of Elderly Singles. Working paper 12554, National Bureau of Economic Research.

De Nardi, M., E. French, and J. B. Jones (2009). Life Expectancy and Old Age Savings. *American Economic Review: Papers and Proceedings 99*, 110–115.

De Nardi, M., E. French, and J. B. Jones (2010). Why Do the Elderly Save? The Role of Medical Expenses. *Journal of Political Economy 118*, 39–75.

Diamond, P. (1998). Optimal Income Taxation: An Example with a U-shaped Pattern of Optimal Marginal Tax Rates. *American Economic Review 88*(11), 83–95.

Diamond, P. A. and J. A. Mirrlees (1971). Optimal Taxation and Public Production I: Production Efficiency. *American Economic Review 61*(1), 8–27.

Diamond, P. A. and E. Saez (2011). The Case for a Progressive Tax: From Basic Research to Policy Recommendations. *Journal of Economic Perspectives 25*(4), 165–90.

Einav, L., A. Finkelstein, and P. Schrimpf (2010). Optimal Mandates and the Welfare Cost of Asymmetric Information: Evidence from the U.K. Annuity Market. *Econometrica 78*, 1031–1092.

Eyraud, L. and L. Lusinyan (2013). Vertical Fiscal Imbalances and Fiscal Performance in Advanced Economies. *Journal of Monetary Economics 60*(5), 571–587.

Farhi, E. and I. Werning (2006). Progressive Estate Taxation. Working paper 12600, National Bureau of Economic Research.

Farhi, E. and I. Werning (2007). Inequality and Social Discounting. *Journal of Political Economy 115*, 365–402.

Farhi, E. and I. Werning (2010). Progressive Estate Taxation. *Quarterly Journal of Economics 125*, 635–673.

Farhi, E. and I. Werning (2013a). Estate Taxation with Altruism Heterogeneity. *American Economic Review: Papers & Proceedings 103*, 489–495.

Farhi, E. and I. Werning (2013b). Insurance and Taxation over the Life Cycle. *The Review of Economic Studies 80*(2), 596–635.

French, E. (2005). The Effects of Health, Wealth, and Wages on Labour Supply and Retirement Behaviour. *Review of Economic Studies 72*(2), 395–427.

Friedman, M. (1972). Second Lecture. In W. Cohen and M. Friedman (Eds.), *Social Security: Universal or Selective?*, pp. 21–49. American Enterprise Institute.

Frydman, C. and R. E. Saks (2010). Executive Compensation: A New View from a Long-Term Perspective, 1936–2005. *Review of Financial Studies 23*(5), 2099–2138.

Gabaix, X. and A. Landier (2008). Why Has CEO Pay Increased So Much? *The Quarterly Journal of Economics 123*(1), 49–100.

Gadenne, L. and M. Singhal (2014). Decentralization in Developing Economies. *Annual Review of Economics 6*, 581–604.

Gale, W. G., J. R. J. Hines, and J. Slemrod (2001). *Rethinking Estate and Gift Taxation*. Brookings Institution Press.

Garicano, L., C. Lelarge, and J. Van Reenen (2013). Firm Size Distortions and the Productivity Distribution: Evidence from France. Working paper.

Garret, D. M. (1995). The Effects of Differential Mortality Rates on the Progressivity of Social Security. *Economic Inquiry 33*, 457–475.

Golosov, M., A. Shourideh, M. Troshkin, and A. Tsyvinsky (2013). Optimal Pension Systems with Simple Instruments. *American Economic Review: Papers & Proceedings 103*, 502–507.

Golosov, M., M. Troshkin, and A. Tsyvinski (2013). Redistribution and Social Insurance. Working paper.

Golosov, M., M. Troshkin, A. Tsyvinski, and M. Weinzierl (2013). Preference Heterogeneity and Optimal Capital Income Taxation. *Journal of Public Economics 97*, 160–175.

Golosov, M., A. Tsyvinski, and I. Werning (2007). New Dynamic Public Finance: A User's Guide. In *NBER Macroeconomics Annual 2006, Volume 21*, pp. 317–388. MIT Press.

Goodspeed, T. (2002). Bailouts in a Federation. *International Tax and Public Finance 9*, 409–421.

Guccio, C., G. Pignataro, and I. Rizzo (2014). Do Local Governments Do it Better? Analysis of Time Performance in the Execution of Public Works. *European Journal of Political Economy 34*, 237–252.

Guesnerie, R. (1981). On Taxation and Incentives: Further Reflections on the Limits to Redistribution. Working paper.

Guner, N., G. Ventura, and Y. Xu (2008). Macroeconomic Implications of Size-Dependent Policies. *Review of Economic Dynamics 11*(4), 721–744.

Hammond, P. (1988). altruism. *The New Palgrave: A Dictionary of Economics First Edition*, 233–239.

Hanson, J. and R. Sigman (2013). Leviathan's Latent Dimensions: Measuring State Capacity for Comparative Political Research. Working paper.

Harsanyi, J. (1995). A Theory of Social Values and a Rule Utilitarian Theory of Morality. *Social Choice and Welfare 12*, 319–344.

Hatfield, J. and G. Padró i Miquel (2012). A Political Economy Theory of Partial Decentralization. *Journal of the European Economic Association 10*(3), 605–633.

Holtz-Eakin, D., D. Joulfaian, and H. S. Rosen (1993). The Carnegie Conjecture: Some Empirical Evidence. *Quarterly Journal of Economics 108*, 413–435.

Hosseini, R. (2014). Adverse Selection in the Annuity Market and the Role for Social Security. *Journal of Political Economy, forthcoming*.

Hosseini, R. and A. Shourideh (2014). Differential Mortality and Optimal Income Taxation. Working paper.

Hurd, M. D., D. McFadden, and A. Merrill (2001). Predictors of Mortality Among the Elderly. Working paper 7440, National Bureau of Economic Research.

Ijiri, Y. and H. A. Simon (1964). Business Firm Growth and Size. *The American Economic Review 54*(2), 77–89.

Inman, R. (2003). Transfers and Bailouts: Enforcing Local Fiscal Discipline with Lessons from U.S. Federalism. In J. Rodden, G. Eskeland, and J. Litvack (Eds.), *Fiscal Decentralization and the Challenge of Hard Budget Constraints*, pp. 35–84. The MIT Press.

Janeba, E. and J. Wilson (2011). Optimal Fiscal Federalism in the Presence of Tax Competition. *Journal of Public Economics 95*(11), 1032–1311.

Joanis, M. (2014). Shared Accountability and Partial Decentralization in Local Public Good Provision. *Journal of Development Economics 107*, 28–37.

Kapička, M. (2013a). Efficient Allocations in Dynamic Private Information Economies with Persistent Shocks: A First-Order Approach. *The Review of Economic Studies 80*(3), 1027–1054.

Kapička, M. (2013b). Efficient Allocations in Dynamic Private Information Economies with Persistent Shocks: A First-Order Approach. *The Review of Economic Studies 80*(3), 1027–1054.

Kaplan, S. N. and J. Rauh (2013). It's the Market: The Broad-Based Rise in the Return to Top Talent. *Journal of Economic Perspectives 27*(3), 35–56.

Katz, L. F. and K. M. Murphy (1992). Changes in Relative Wages, 1963-1987: Supply and Demand Factors. *The Quarterly Journal of Economics 107*(1), 35–78.

King, R. G. and S. T. Rebelo (1999). Resuscitating Real Business Cycles. *Handbook of Macroeconomics 1*, 927–1007.

Kitagawa, E. M. and P. M. Hauser (1973). *Differential Mortality in the United States: A Study in Socioeconomic Epidemiology*. Harvard University Press.

Kocherlakota, N. R. (2005). Zero Expected Wealth Taxes: A Mirrlees Approach to Dynamic Optimal Taxation. *Econometrica 73*, 1587–1621.

Kocherlakota, N. R. (2010). *The New Dynamic Public Finance*. Princeton University Press.

Kornai, J., E. Maskin, and G. Roland (2004). Understanding the Soft Budget Constraint. *Journal of Economic Literature 41*(4), 1095–1136.

Köthenbürger, M. (2004). Tax Competition in a Fiscal Union with Decentralized Leadership. *Journal of Urban Economics 55*(3), 498–513.

Köthenbürger, M. (2011). How Do Local Governments Decide on Public Policy in Fiscal Federalism? Tax vs. Expenditure Optimization. *Journal of Public Economics 95*(12), 1516–1522.

Kotlikoff, L. J. (1988). Intergenerational Transfers and Savings. *Journal of Economic Perspectives 2*, 41–58.

Kotlikoff, L. J. and L. Summers (1981). The Role of Intergenerational Transfers in Aggregate Capital Accumulation. *Journal of Political Economy 89*, 706–732.

Krueger, D. and F. Perri (2006). Does Income Inequality Lead to Consumption Inequality? Evidence and Theory. *Review of Economic Studies 73*, 163–193.

Lazear, E. P., K. L. Shaw, and C. Stanton (2013). Making Do With Less: Working Harder During Recessions. Working paper 19328, National Bureau of Economic Research.

Lewellen, W. G. and B. Huntsman (1970). Managerial Pay and Corporate Performance. *The American Economic Review 60*(4), 710–720.

Litvack, J., J. Ahmad, and R. Bird (1998). *Rethinking Decentralization in Developing Countries*. The World Bank.

Loayza, N., J. Rigolini, and O. Calvo-González (2014). More Than You Can Handle: Decentralization and Spending Ability of Peruvian Municipalities. *Economics and Politics 26*(1), 56–78.

Lucas, R. E. (1978). On the Size Distribution of Business Firms. *Bell Journal of Economics 9*(2), 508–523.

MaCurdy, T. E. (1981). An Empirical Model of Labor Supply in a Life-Cycle Setting. *The Journal of Political Economy 89*(6), 1059–1085.

Malik, H. J. (1970). Estimation of the Parameters of the Pareto Distribution. *Metrika 15*(1), 126–132.

Mankiw, G., M. Weinzierl, and D. Yagan (2009). Optimal Taxation in Theory and Practice. *Journal of Economic Perspectives 23*(4), 147–174.

Mann, M. (1984). The Autonomous Power of the State: Its Origins, Mechanisms and Results. *Archives Européennes de Sociologie 25*(2), 185–213.

Mayer, T. (1960). The Distributions of Ability and Earnings. *The Review of Economics and Statistics 42*(2), 189–195.

Mirrlees, J. A. (1971). An Exploration in the Theory of Optimum Income Taxation. *The Review of Economic Studies 38*(2), 175–208.

Modigliani, F. (1988). The Role of Intergenerational Transfers and Life Cycle Saving in the Accumulation of Wealth. *Journal of Economic Perspectives 2*, 15–40.

Oates, W. (2005). Toward a Second-Generation Theory of Fiscal Federalism. *International Tax and Public Finance 12*(4), 349–373.

Panis, C. and L. Lillard (1995). Socioeconomic Differentials in the Returns to Social Security. Working paper.

Patil, S., A. Gupta, D. Desai, and A. Sajane (2013). Causes of Delay in Indian Transportation Infrastructure Projects. *International Journal of Research in Engineering and Technology 2*(1), 71–80.

Pavan, A., I. Segal, and J. Toikka (2014). Dynamic Mechanism Design: A Myersonian Approach. *Econometrica 82*(2), 601–653.

Peralta, S. (2011). Partial Fiscal Decentralization, Local Elections and Accountability. Working paper.

Persson, T. and G. Tabellini (1992). The Politics of 1992: Fiscal Policy and European Integration. *Review of Economic Studies 59*(4), 689–710.

Pettersson-Lidbom, P. (2010). Dynamic Commitment and the Soft budget Constraint: An Empirical Test. *American Economic Journal: Economic Policy 2*(3), 154–179.

Phelan, C. (2006). Opportunity and Social Mobility. *Review of Economic Studies 73*, 487–505.

Pijoan-Mas, J. and J.-V. Ríos-Rull (2014). Heterogeneity in Expected Longevities. *Demography, forthcoming*.

Piketty, T. (2014). *Capital in the Twenty-First Century*. Harvard University Press.

Piketty, T. and E. Saez (2003). Income Inequality in the United States, 1913-1998. *The Quarterly Journal of Economics 118*(1), 1–39.

Piketty, T. and E. Saez (2013). A Theory of Optimal Inheritance Taxation. *Econometrica 81*, 1851–1886.

Piketty, T., E. Saez, and S. Stantcheva (2014). Optimal Taxation of Top Labor Incomes: A Tale of Three Elasticities. *American Economic Journal: Economic Policy 6*(1), 230–71.

Prescott, E. C. (2004). Why do Americans Work So Much More than Europeans? *Federal Reserve Bank of Minneapolis Quarterly Review 28*(1), 2–13.

Prud'homme, R. (1995). The Dangers of Decentralization. *World Bank Research Observer 10*(3), 201–220.

Qian, Y. and G. Roland (1998). Federalism and the Soft-Budget Constraint. *American Economic Review 88*(5), 1146–1162.

Quadrini, V. (2000). Entrepreneurship, Saving and Social Mobility. *Review of Economic Dynamics 3*(1), 1–40.

Rand, A. (1957). *Atlas Shrugged*. Penguin.

Roberts, D. R. (1956). A General Theory of Executive Compensation Based on Statistically Tested Propositions. *The Quarterly Journal of Economics 70*(2), 270–294.

Rodden, J., G. Eskeland, and J. Litvack (2003). *Fiscal Decentralization and the Challenge of Hard Budget Constraints*. The MIT Press.

Rosen, S. (1982). Authority, Control, and the Distribution of Earnings. *The Bell Journal of Economics 13*(2), 311–323.

Rothschild, C. and F. Scheuer (2013). Redistributive Taxation in the Roy Model. *The Quarterly Journal of Economics 128*(2), 623–668.

Sadka, E. (1976). On Income Distribution, Incentive Effects and Optimal Income Taxation. *Review of Economic Studies 43*(1), 261–268.

Saez, E. (2001). Using Elasticities to Derive Optimal Income Tax Rates. *The Review of Economic Studies 68*(1), 205–229.

Saez, E. (2002). The Desirability of Commodity Taxation under Non-Linear Income Taxation and Heterogeneous Tastes. *Journal of Public Economics 83*, 217–230.

Saez, E. (2003). The Effect of Marginal Tax Rates on Income: A panel Study of 'Bracket Creep'. *Journal of Public Economics 87*(5-6), 1231–1258.

Saez, E., J. Slemrod, and S. H. Giertz (2012). The Elasticity of Taxable Income with Respect to Marginal Tax Rates: A Critical Review. *Journal of Economic Literature 50*(1), 3–50.

Sanguinetti, P. and M. Tommasi (2004). Intergovernmental Transfers and Fiscal Behavior Insurance Versus Aggregate Discipline. *Journal of International Economics 62*(1), 149–170.

Scheuer, F. (2014). Entrepreneurial Taxation with Endogenous Entry. *American Economic Journal: Economic Policy 6*(2), 126–163.

Shourideh, A. (2012). Optimal Taxation of Entrepreneurial Income: A Mirrleesian Approach to Capital Accumulation. Working paper.

Shourideh, A. and M. Troshkin (2012). Providing Efficient Incentives to Work: Retirement Ages and the Pensioin System. Working paper.

Silva, E. and A. Caplan (1997). Transboundary Pollution Control in Federal Systems. *Journal of Environmental Economics and Management 34*(2), 173–186.

Simon, H. A. and C. P. Bonini (1958). The Size Distribution of Business Firms. *The American Economic Review 48*(4), 607–617.

Singh, G. K. and M. Siahpush (2006). Widening Socioeconomic Inequalities in U.S. Life Expectancy, 1980-2000. *International Journal of Epidemiology 35*, 969–979.

Slavík, C. and H. Yazıcı (2014). Machines, Buildings and Optimal Dynamic Taxes. Forthcoming, *Journal of Monetary Economics*.

Slemrod, J. (2000). *Does Atlas Shrug?: The Economic Consequences of Taxing the Rich*. Harvard University Press.

Slemrod, J. and V. Venkatesh (2002). The Income Tax Compliance Cost of Large and Mid-Size Businesses. Working paper.

Snyder, R. (2001). Scaling Down: The Subnational Comparative Method. *Studies in Comparative International Development 36*(1), 93–110.

Spear, S. E. and S. Srivastava (1987). On Repeated Moral Hazard with Discounting. *Review of Economic Studies 54*, 599–617.

Steiner, S. (2010). How Important is the Capacity of Local Governments for Improvements in Welfare? Evidence from Decentralised Uganda. *Journal of Development Studies 46*(4), 644–661.

Stiglitz, J. E. (1982). Self-Selection and Pareto Efficient Taxation. *Journal of Public Economics 17*(2), 213–240.

Tomes, N. (1981). The Family, Inheritance, and the Intergenerational Tansmission of Inequality. *Journal of Political Economy 89*, 928–958.

Vaupel, J. W., K. G. Manton, and E. Stallard (1979). The Impact of Heterogeneity in Individual Frailty on the Dynamics of Mortality. *Demography 16*, 439–454.

Waldron, H. (2007). Trends in Mortality Differentials and Life Expectancy for Male Social Security-Covered Workers, by socioeconomic status. *Social Security Bulletin 67*, 1–28.

Weingast, B. (2009). Second Generation Fiscal Federalism: The Implications of Fiscal Incentives. *Journal of Urban Economics 65*(3), 279–293.

Wildasin, D. (1988). Nash Equilibria in Models of Fiscal Competition. *Journal of Public Economics 35*(2), 229–240.

Wildasin, D. (1997). Externalities and Bailouts Hard and Soft Budget Constraints in Intergovernmental Fiscal Relations. Working paper.

Wilson, J. (1991). Tax Competition with Interregional Differences in Factor Endowments. *Regional Science and Urban Economics 21*(3), 423–451.

Wooldridge, J. M. (2002). *Econometric Analysis of Cross Section and Panel Data*. MIT press.

Ziblatt, D. (2008). Why Some Cities Provide More Public Goods than Others: A Subnational Comparison of the Provision of Public Goods in German Cities in 1912. *Studies in Comparative International Development 43*(3-4), 273–289.

Zodrow, G. and P. Mieszkowski (1986). Pigou, Tiebout, Property Taxation, and the Underprovision of Local Public Goods. *Journal of Urban Economics 19*(3), 356–370.