Three Essays on Public Economics and Heterogeneity

A DISSERTATION

SUBMITTED TO THE FACULTY OF THE GRADUATE SCHOOL

OF THE UNIVERSITY OF MINNESOTA

BY

Anderson Luís Schneider

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

Larry Jones

August, 2009

Acknowledgments

I am greatly indebted to Larry Jones, Narayana Kocherlakota, Christopher Phelan and V.V. Chari for their continuous help, guidance and support. I would like to thank Roozbeh Hosseini, Francesca Carapella, Laurence Ales, Pricila Maziero, Daniel Samano, Hakki Yazici, Adam Slawski, Lei Guo, John Dalton, Erzo Luttmer, Ellen McGrattan, Pierre Yared and José Víctor Ríos Rull for comments and suggestions. I am also grateful to Art Rolnick and Warren Weber for allowing me to spend an invaluable time at the Minneapolis Fed.

Great appreciation is reserved to Facundo Piguillem for the friendship and inumerous discussions. This thesis is based on joint work with Facundo.

My greatest gratitude goes to my parents, my sister and Pricila for the unconditional love and support.

Dedication

This dissertation is dedicated to my family.

Contents

List of Tables

List of Figures

Part I

Introduction

This thesis presents a collection of essays about Public Economics and individual heterogeneity. The essays are motivated by two different subjects. The first subject refers to the relation between economic outcomes and majority voting in a democratic regime. More specifically, the outcome regarding redistributive labor income taxes is analyzed when heterogeneous individuals vote, once and for all, over an infinite sequence of taxes. The second subject refers to time consistency problems in Public Economics. Issues about optimal fiscal policy are considered in an environment where different individuals hold distinct information about (sequential) action performed by the government. This friction prevents the standard punishment mechanism that enforces good policy outcomes or, alternatively, inhibits the occurrence of time consistency problems. The best equilibrium outcome is then analyzed in this new situation.

Chapters 2 focus on the first subject. Moreover, the chapter explores the relationship between changes in labor income inequality and movements in labor taxes over the last decades in the US. In order to do so, this relation is modeled through a political economy channel by developing a median voter result over sequence of taxes. We consider an infinite horizon economy in which agents are heterogeneous with respect to both initial wealth and labor skills. We study indirect preferences over redistributive fiscal policies - sequences of affine taxes on labor and capital income - that can be supported as a competitive equilibrium. The paper assumes balanced growth preferences and full commitment. The first result is the following: if initial capital holdings are an affine function of skills, then the best fiscal policy for the agent with the median labor skill is preferred to any other policy by at least half of the individuals in the economy. The second result provides the characterization of the most preferred tax sequence by the median agent: marginal taxes on labor depend directly on the absolute value of the distance between the median and the mean value of the skill distribution. We extend the above results to an economy in which the distribution of skills evolves stochastically over time. A temporary increase in inequality could imply either higher or lower labor taxes, depending on the sign of the correlation between inequality and aggregate labor. The calibrated model does a good job on fitting both the increasing trend and the levels of labor taxes in the last decades, and also on matching some short run co-movements.

Chapter 3 generalizes the median voter theorem developed in chapter 2 to a situation where there is no commitment or, alternatively, voting is sequential over time. More specifically, the same equilibrium definition as in Bernheim and Slavov (2008) is adopted.

Chapter 4 deals with optimal fiscal policy when the government takes actions sequentially over time and cannot commit to a pre-specified plan of actions. These features potentially generate what is known in the literature as time consistency problems. Although these problems play an important role in public policy, game theoretical models in macroeconomics seem to indicate the opposite. Due to the complexity of this kind of models, it is commonly assumed that information is complete and perfect. In turn, this assumption becomes the key element that allows agents to coordinate perfectly to punish the government if it does not do what private agents want. As a result, a wide range of feasible payoffs can be sustained as equilibrium, including the best payoff under commitment. Since this approach is widely used for normative purposes a natural question emerges: are the above results robust to small variations in information? This paper analyzes an investment taxation problem in an economy with incomplete information. Specifically, we study an environment with the following main characteristics: 1) the aggregate productivity (fundamental) is stochastic, 2) only the government observes it and; 3) every agent privately receives a noisy signal about the fundamental. The first characteristic implies that the best policy (tax on investment) with commitment is state contingent. The second and third characteristics make the information incomplete. In particular, agents have different information sets, and therefore different beliefs, about the true state of the economy. As a result, independently of the accuracy of the signal, incomplete information reduces the set of equilibrium payoffs. First, we show that any policy that depends solely on the fundamental cannot be an equilibrium. Second, the best equilibrium policy is independent of the fundamental. Finally, for any discount factor strictly smaller than one and for any size of the noise, the best equilibrium is inefficient.

Part II

Heterogeneous Labor Skills, The Median Voter and Labor Taxes[1]

1 Introduction

Even though a great extent of work has been done to analyze the characteristics of taxes determined by politico-economic process, little is known about the properties of labor taxes. One way to construct a positive theory about fiscal policy is to assume that agents optimally choose the policies that will take place in the future. This optimal choice may, in part, be motivated by a first order issue in elections: the redistributive effects of the different policies. Following this idea, we study a class of dynamic model economies with heterogeneous agents where the only political institution is the pursuit of consensus. Agents vote, once and for all, at the beginning of time on sequences of redistributive taxes on capital and labor. Building on the work of Bassetto and Benhabib (2006) (henceforth B&B), we derive a median voter theorem for this class of economies. We use this theorem to describe the properties of the equilibrium tax sequences. The theorem gives one, precise, statement of the form that redistribution considerations take in determining policy.

As in B&B, we add fiscal policy that allows for redistribution in the standard

[1] This chapter is coauthored with **Facundo Piguillem.**

neoclassical growth model. We go beyond their paper by adding leisure choice and stochastically evolving labor productivities. Along with this, we also add to their framework marginal taxes on labor income. Individual's disagreement over capital income taxes is motivated by differences in initial wealth levels. The conflict about labor taxes is given by heterogeneous labor skills: agents have different abilities to turn effort into effective labor, which is rented to firms in the market. Although we do not model any voting process explicitly, consider the following situation: at time zero, before the economy starts, all possible sequences representing different fiscal policies are analyzed, and a consensus should be reached through sincere majority voting. By sincere we mean an agent would vote for fiscal policy A against policy B if she prefers A to B. A natural question then arises: is it actually possible to reach such a consensus? Or in other words, is there any policy that precludes the existence of Condorcet cycles when sincere voting is in place? If such a policy does exist, what are the capital and labor income taxes implied and how does individual heterogeneity shape this policy?

In proposition 1, assuming balanced growth preferences, we give two different sets of sufficient conditions for the existence of a Condorcet winner in this type of environment. The first part of proposition 1 assumes that there is no heterogeneity in labor skills, although agents value leisure. In this case, we show that, independent of the distribution of initial endowments, the best fiscal policy (consisting of a full time path of both labor and capital income tax rates) for the agent with the median endowment is preferred to any other policy by at least half of the individuals in the economy. The second part of the proposition considers heterogeneity in both labor skills and initial wealth. If the initial capital endowment is an affine function of labor skills, then again, the most preferred time path of policies is preferred by a majority. The proof of the consensus result relies on a characterization of indirect preferences over fiscal policies that is of independent interest. Although proposition 1 can be thought in terms of fiscal policies, it is actually stated in terms of implementable allocations: those that can be decentralized as a competitive equilibrium. Under complete markets, if agents have the same balanced growth utility function, individual allocations can be expressed as a constant share of their aggregate counterparts. These shares are functions of both types and aggregate allocations. Moreover, the indirect preferences, as a function of types, inherit the properties of these share functions. Then, we show that for any two fiscal policies for which a competitive equilibrium exists, the indirect preferences can cross at most once in the space of types, delivering the result.

Our second contribution concerns the characterization of the Condorcet Winner. It follows that the indirect utility for the median type over any fiscal policy can be decomposed into two parts: a redistributive component and an efficiency one. The redistributive component depends directly on the skewness of the distribution of skills and it is increasing in the distortions yielded by both capital and labor taxes (together with bigger transfers). The efficiency component is given by the value of the mean type's indirect utility and it is decreasing in any positive distortion. The most preferred tax schedule for the median type balances these two components. As in B&B, we show that capital income taxes will be either zero or at the upper bound in any period and state, with at most one period in between.

In addition, Proposition 3 shows that marginal taxes on labor income depend directly on the absolute value of the distance between the median and the mean value of the productivity distribution in the economy.

The results are extended to the case that skills evolve stochastically over time keeping constant the ranking among agents. Again labor taxes depend directly on the skewness of the distribution. However, the final outcome is ambiguous - this relationship could be either increasing or decreasing. We study a numerical example and show that the model without capital accumulation can generate either procyclical or counter-cyclical taxes, depending on how the dispersion of the distribution of individual productivities changes along the business cycle. This may offer an answer to the question posed by Alesina, Campante, and Tabellini (2008), related to the empirical observation that fiscal policy is often procyclical in developing countries and counter-cyclical in developed ones.

Finally, a calibrated version of the model without initial wealth inequality is used to check if the theory can account for the observed increasing trend in both labor income inequality and average tax on labor in US in the last decades. The calibration for the skill process is done using data on wages from Eckstein and

Nagypal (2004). The model does a good job on fitting both the increasing trend and the levels of labor taxes in the last decades, and also on matching some short run co-movements. The model accounts for twice as much of the growth in labor taxes observed in the period 1962-2001. It also yields a negative correlation between taxes and aggregate labor, in line with the data.

We view the results regarding labor taxes as a neat characterization of an important component of fiscal policy. Most of the work that analyzes inefficiencies due to political constraints, has followed the route of making strong, ex ante, assumptions about the forms of institutions. The difficulty with this approach is that it typically requires making very specific assumptions about the institutions that are used to generate policies (e.g., specific game theoretic models of voting over a restricted set of tax instruments) along with a variety of other imperfections. This makes it hard to interpret the results since it is not clear if the properties of the policies singled out as equilibria are chosen due to the specific institutional arrangements assumed or due to the imperfections added.

As we mentioned before, our model gives an extension to the median voter result presented in Bassetto and Benhabib (2006).[2] Other than the already highlighted differences in the characteristics of the physical environment, our results depend on the assumption of balanced growth preferences defined over consumption and

[2]Important contributions on median voter results and its connection with fiscal policy include Meltzer and Richard (1981), Alesina and Rodrik (1994), Persson and Tabellini (1994), among others.

leisure. On the other hand, although B&B do not consider leisure choice, a more general class of Gorman aggregable preferences is analyzed.

Besides the work by Bassetto and Benhabib (2006), five other papers deserve special mention. Werning (2007) considers the same physical environment as here and analyzes the Ramsey outcome when the government uses fiscal policy for redistribution and to finance an exogenous stream of expenditures. The possibility of non distortionary taxation is not ruled out ex-ante, nevertheless distortions emerge in the decentralized solution, regardless of the welfare weights used by the government. We find similar results, although we obtain a more specific characterization of labor taxes. This feature comes partially from the fact that the median voter solution uses welfare weight equal to one for the median type. In the case that agents have stochastic labor skills, a numerical exercise in his paper shows that the implied labor taxes from a Utilitarian Ramsey problem comove with the distribution of skills. The author does not provide a numerical solution calibrated to the US economy.

Azzimonti, de Francisco, and Krusell (2008) also analyze majority voting over marginal taxes on labor income. Since their environment does no consider both aggregate uncertainty and capital accumulation, the best sequence of labor taxes for each type in the economy can be characterized by two numbers (taxes in the first two periods). A median voter result is provided in the case where either there is heterogeneity in the initial wealth only or in the labor skills.

Krusell and Rios-Rull (1999) consider an environment similar to ours but voting takes place periodically, taxes on capital and labor income are constrained to be equal and only future taxes can be changed. A Markov stationary equilibrium is solved numerically. The stationary equilibrium exhibits positive distortions. As in this paper, the level of income taxation depends on the skewness of income distribution. Since their paper consider a marginal tax on income, results about labor taxes are not provided.[3]

Regarding the empirical results, Chari, Christiano, and Kehoe (1994) analyze the quantitative implications of optimal fiscal policy in a dynamic model with homogenous agents. We emphasize that they consider the same class of balanced growth preferences. Using different calibrated versions of the model, they found that labor taxes are essentially constant over the business cycle, although labor taxes inherit the stochastic properties of the exogenous shocks (productivity and government spending). Finally, Corbae, D'Erasmo, and Kuruscu (2008) use a recursive political economy model, as in Krusell and Rios-Rull (1999), to evaluate how much the increase in wage inequality in the period 1979-1996 can account for the relative increase in both transfers to low earnings quintiles and effective tax rates for higher quintiles. They assume idiosyncratic labor skills shocks and incomplete markets. The paper uses a median voter result by checking numerically that preferences over one period income taxes (on and off-path) are single-peaked. They found

[3] Azzimonti, de Francisco, and Krusell (2006) provide an analytical characterization of time-consistent Markov-perfect equilibria in an environment similar to Krusell and Rios-Rull (1999), but individual heterogeneity is restricted to initial wealth.

that the model predicts about half of the increase in redistribution to lowest wage quintiles, and also it overpredicts the average effective tax rate.

The paper proceeds as follows. Section II describes the environment. Section III characterizes the competitive equilibrium given a fiscal policy. In section IV we construct the proof for the consensus result. Section V characterizes the Condorcet winner, while section VI considers stochastic skills. Section VII shows the numerical results and the last section concludes.

2 The Economy with Constant Skills

There is a continuum of agents indexed by the labor skill parameter $\theta \in \Theta \equiv [\underline{\theta}, \overline{\theta}]$ with $\underline{\theta} > 0$. Later we relax this assumption, allowing for stochastic labor skills. The distribution of θ is represented by the p.d.f. $f(\cdot)$ and the median type is denoted by $\theta^m \leq \int_\Theta \theta f(\theta) d\theta = 1$ by assumption.[4]

Uncertainty is driven by the public observable state $s_t \in S$, where S is finite. It potentially affects the efficient production frontier. Let $s^t = (s_0, ..., s_t)$ be the history of shocks up to time t and $\Pr(s^t)$ its marginal probability. We assume that $\Pr(s_0 = \overline{s}) = 1$ for some $\overline{s} \in S$.

The output at time t is produced by competitive firms using capital and efficient labor. The resources constraint for each pair (t, s^t) is

[4] The median voter result presented later does not depend on this skewness assumption.

$$C(s^t) + K(s^t) \leq F\Big(L(s^t), K(s^{t-1}), s^t\Big) + (1-\delta)K(s^{t-1}) \tag{1}$$

where the function $F(\cdot)$ is assumed to be homogeneous of degree one in both capital and labor for all s^t.

Remark: We could consider an exogenous stream of government expenditures without changing the main results. But since our concern is mainly related to redistribution, the restriction of zero government consumption avoids dealing with valuations of the benefits of positive marginal taxes net of the distortions in financing government expenditures.

Each agent has an endowment of one unit of time in each period and state. Using l/θ units of its time agent type θ produces l units of efficient labor that is rented to the firms. If agent type θ consumes the stream $\{c_t, 1-l_t/\theta\}_{t=0}^{\infty}$ of consumption and leisure, then its total discounted utility is given by $\sum_{t=0}^{\infty} \beta^t u\Big(c(s^t,\theta), 1-l(s^t,\theta)/\theta)\Big)$, where:

$$u(c, l_e) = \begin{cases} \frac{\left[c^{\alpha} l_e^{1-\alpha}\right]^{1-\sigma}}{1-\sigma} & \text{if } \sigma \neq 1 \\ \alpha \log(c) + (1-\alpha)\log(l_e) & \text{if } \sigma = 1 \end{cases} \tag{2}$$

In the initial period, agent type θ is endowed with $k_{-1}(\theta) > 0$ units of capital stock. Later we shall impose conditions on the initial wealth distribution.

In each period the government levies an affine tax schedule on labor income given by $\tau_l(s^t)w(s^t)l(s^t;\theta) + T(s^t)$, where $w(s^t)$ are the wage payments and the lump-sum tax $T(s^t)$ is potentially used for redistribution. Notice that the tax schedule is not individual specific.

The government taxes capital returns net of depreciation at rate $\tau_k(s^t) \in [0, \overline{\tau}]$. For the type of wealth distribution that we analyze later, the lower bound will never bind. The upper bound on capital taxes is a technical condition required in order to guarantee that the best allocation for the median type exists. In order to reduce the arbitrariness of such an exogenous upper bound, we choose $\overline{\tau} = 100\%$. In this way the maximum levy corresponds to a loss of the full return net of depreciation.

Profit maximization by the firms determines the rental prices. Given a tax sequence, prices, and initial endowments, under complete markets agent type θ chooses his individual allocation in order to maximize utility subject to the budget constraint:

$$\sum_{t,s^t} p(s^t)\Big(c(s^t;\theta)+k(s^t;\theta)\Big) \leq \sum_{t,s^t} p(s^t)\Big((1-\tau_l(s^t))w_t(s^t)l(s^t;\theta)+R(s^t)k(s^{t-1};\theta)\Big)-T \tag{3}$$

where $T \equiv \sum_{t,s^t} p(s^t)T(s^t)$ is the present value of the lump-sum taxes and $R(s^t) \equiv 1 + (1 - \tau_k(s^t))(r(s^t) - \delta)$.

Under the complete markets assumption the government budget constraint can be written as:

$$-T \leq \sum_{t,s^t} p(s^t)\Big(\tau_l(s^t)w(s^t)L(s^t) + \tau_k(s^t)(r(s^t) - \delta)K(s^{t-1})\Big) \tag{4}$$

The usual definition for a competitive equilibrium follows:

Definition 1. *A competitive equilibrium given taxes* $\{\tau_l(s^t), \tau_k(s^t), T(s^t)\}_{t=0}^{\infty}$ *is a sequence of prices* $\{w(s^t), p(s^t), r(s^t)\}_{t=0}^{\infty}$*, individual allocations* $\{c(s^t;\theta), l(s^t;\theta), k(s^t;\theta)\}_{t=0}^{\infty}$ *and implied aggregate allocations* $\{C(s^t), L(s^t), K(s^t)\}_{t=0}^{\infty}$ *such that:*

1. *Given after-tax prices,* $\{c(s^t;\theta), l(s^t;\theta), k(s^t;\theta)\}_t$ *maximizes utility subject to (3);*

2. $C(s^t) = \int_\Theta c(s^t;\theta)f(\theta)d\theta$, $L(s^t) = \int_\Theta l(s^t;\theta)f(\theta)d\theta$, *and* $K(s^t) = \int_\Theta k(s^t;\theta)f(\theta)d\theta$*;*

3. *Factor prices are equal to the marginal products for every* s^t*;*

4. *The government budget constraint holds for every* s^t*; and*

5. *The resource constraint holds for every* s^t*.*

3 Equilibrium Characterization

Here we characterize the economy given a fiscal policy for the log utility case.[5] We use a characterization strategy similar to Werning (2007).

Let $\lambda(\theta)$ be the multiplier related to the budget constraint of type θ. The first

[5] It turns out that the characterization in the logarithmic case is much simpler than in the general case (balanced growth preferences). All the proofs in the general case are shown in the appendix 1.

order conditions with respect to individual consumption and output yield:

$$\begin{aligned}
\frac{\alpha\beta^t \Pr(s^t|s_0)}{c(s^t;\theta)} &= p(s^t)\lambda(\theta) && (5)\\
\frac{(1-\alpha)\beta^t \Pr(s^t|s_0)}{\theta - l(s^t;\theta)} &= p(s^t)(1-\tau_l(s^t))w_t(s^t)\lambda(\theta) && (6)
\end{aligned}$$

Let $\varphi(\theta) \equiv 1/\lambda(\theta)$, and $E(\varphi) \equiv \int_\Theta \varphi(\theta) f(\theta) d\theta$. Then integration over types in the expressions above yields the following equations determining after-tax prices:

$$\begin{aligned}
p(s^t) &= \frac{E(\varphi)\alpha\beta^t \Pr(s^t|s_0)}{C(s^t)} && (7)\\
p(s^t)(1-\tau_l(s^t))w_t(s^t) &= \frac{E(\varphi)(1-\alpha)\beta^t \Pr(s^t|s_0)}{1-L(s^t)} && (8)
\end{aligned}$$

We normalize the initial price p_0 so that $E(\varphi) = 1$.[6] Then individual allocations can be written as:

$$\begin{aligned}
c(s^t;\theta) &= \varphi(\theta)C(s^t) && (9)\\
1 - l(s^t;\theta)/\theta &= \varphi(\theta)\theta^{-1}[1-L(s^t)] && (10)
\end{aligned}$$

The other conditions for optimization in the problem faced by individual θ are

$$p(s^t) = \sum_{s^{t+1}} R(s^{t+1})p(s^{t+1}), \quad \text{and} \quad \lim_{t\to\infty}\sum_{s^t} p(s^t)k(s^t;\theta) = 0$$

Using conditions (5)-(9) in each individual's budget constraint yield:

$$\varphi(\theta) = (1-\beta)\left[\widetilde{W}_0(\theta, T, \tau_{k0}) + \theta\sum_{t,s^t}\beta^t \Pr(s^t)\left[\frac{(1-\alpha)}{(1-L(s^t))}\right]\right] \qquad (11)$$

[6] In the proof of Lemma 1 we show that individual shares integrate to one for any normalization of initial prices.

where $\widetilde{W}_0(\theta, T, \tau_0) \equiv (\alpha/C_0)R_0 k_{-1}(\theta) - T$.

Since for each type the expression for $\varphi(\cdot)$ depends on the aggregate allocations and the tax schedule, the function can be rewritten as $\varphi(Z;\theta) \in \mathbb{R}_+$, where Z is a sequence consisting of aggregate allocations, initial tax on capital and the present value lump-sum transfer. Let Z^∞ be the set of such sequences.

From (11) we have that the share for type θ is equal to the after-tax value of his initial wealth plus the maximal present discounted value of his labor income. Next we give a more intuitive representation of the function φ. Using (11) and the fact that $E(\varphi) = 1$, the individual shares can be rewritten as:

$$\varphi(Z;\theta) = 1 + (1-\beta)\left[\left(\widetilde{W}_0(\theta, T, \tau_{k0}) - \mathrm{E}(\widetilde{W}_0(\theta, T, \tau_{k0}))\right) + (\theta - 1)\cdot U_L(Z)\right] \quad (12)$$

where $U_L(Z) \equiv \sum_{t,s^t} \beta^t \Pr(s^t)\left[\frac{(1-\alpha)}{(1-L(s^t))}\right]$.

Therefore individuals that are wealthier than the average will have both individual consumption and leisure (measured in efficient units) higher than the respective aggregates.

Remark: For future use, it is straightforward to replicate the computations above for the case in which the heterogeneity is only restricted to the initial endowments. In this case we would have individuals indexed by the initial capital endowment distributed according a pdf $f(\cdot)$ on $[\underline{k}, \overline{k}]$. The labor skills are given by the constant function $\theta(k_{-1}) = 1 \; \forall \; k_{-1} \in [\underline{k}, \overline{k}]$. In this case one can find that $\varphi(Z; k_{-1}) = (1-\beta)[\widetilde{W}_0(k_{-1}, T, \tau_0) + \sum_{s^t} \beta^t \Pr(s^t)(1-\alpha)/(1-L(s^t))]$.

Then, similar to Werning (2007), we have the following:

Lemma 1. *$Z \equiv (\{C(s^t), L(s^t), K(s^t)\}_{t\geq 0}, T, \tau_{k0})$ is the aggregate allocation sequence (together with T and $\tau_{k0} \leq \overline{\tau}$) in an interior CE if and only if:*

1. *Z satisfy the resources constraint in (1) for all s^t;*

2. *$\frac{1}{C(s^t)} \geq \beta E\left[\frac{1+(1-\overline{\tau})(F_k(s^{t+1})-\delta)}{C(s^{t+1})}|s^t\right]$ for all s^t;*

3. *Evaluated at the aggregate allocations, the function $\varphi : Z^{\infty} \times \Theta \to \mathbb{R}_+$ given by (12) is such that $\varphi(Z;\theta) \in (0, \frac{1}{1-L(s^t)})$ for all $s^t, \theta \in \Theta$, and $L(s^t) \subset Z$.*

Proof: See the appendix 1.

Necessity comes from the reasoning above. Sufficiency is shown in the appendix 1. The second condition comes from the upper bound on the capital tax rates. The last condition comes from the nonnegativity of consumption and the fact that leisure is bounded by the unit. It replaces the usual implementability conditions found in the Ramsey literature.

Depending on the restrictions on the distribution of the initial endowment, we could relax the third condition in the Lemma above to $\varphi(Z;\theta) \geq 0$. For example, this would be the case if initial endowments are non-decreasing in the skill level, making $\varphi(\cdot)$ strict increasing in θ.

Because preferences are homothetic, Lemma 1 implies that, given taxes, two economies having different distributions of productivity types with the same mean and the same initial aggregate capital stock will have the same aggregate outcomes in equilibrium. Clearly, the distribution of φ in the economy will indeed depend on the distribution of skills and the assumptions on the initial endowments. Also notice that the distortions generated by marginal taxes are enclosed in the aggregates that determine the function φ.

As Chari and Kehoe (1999) have pointed out, the non-arbitrage condition $p(s^t) = \sum_{s_{t+1}} p(s^{t+1})R(s^{t+1})$ does not uniquely pin down the stochastic process for the capital tax rate.

4 The Consensus Result

Let Ξ be the set of elements $Z \equiv (\{C(s^t), L(s^t), K(s^t)\}_{s^t}, T, \tau_{k0})$ that satisfy the conditions in Lemma 1.

We begin by analyzing preference orderings over elements of Ξ. As we have pointed out, all distortions generated by marginal taxes are already enclosed in the aggregate allocations. Given any $Z \in \Xi$, we can express the present discounted utility for each individual when aggregate allocations are given by Z. For agent type θ, denote this value by $V(Z;\theta)$. Then we have:

$$V(Z;\theta) = \frac{1}{1-\beta}\left[\log\Big(\varphi(Z;\theta)\Big) - (1-\alpha)\log(\theta)\right] + \sum_t \beta^t \Pr(s^t)\left[\alpha\log(C(s^t)) + (1-\alpha)\log(1-L(s^t))\right] \quad (13)$$

The share of each individual can be rewritten as $\varphi(Z;\theta) = Bk_{-1}(\theta) + C\theta + D$, where $B, C,$ and D are values that depend on $Z \in \Xi$. For a given $Z \in \Xi$ and associated $\varphi(\cdot)$, with some abuse of notation, let $J_\varphi(\theta)$ and $J(Z)$ be respectively the first and second term of (13).

Agent θ weakly prefers the allocation Z to $\hat{Z}$ if and only if

$$V(Z;\theta) \geq V(\hat{Z};\theta) \Longleftrightarrow J_{\varphi}(\theta) + J(Z) \geq J_{\hat{\varphi}}(\theta) + J(\hat{Z})$$

Let $S_{Z,\hat{Z}} = \left\{\theta : J_{\varphi}(\theta) - J_{\hat{\varphi}}(\theta) \geq J(\hat{Z}) - J(Z)\right\}$, that is, the set of agents that prefers Z to $\hat{Z}$.

As mentioned before, in this section we state the results for the logarithmic case. The proof for the more general class of utility functions is shown in the appendix 1. The general strategy of the proof presented below is similar to the one used to prove proposition 2 in Benhabib and Przeworski (2006).

Proposition 1. *Assume balanced growth preferences in (2) and consider any $Z, \hat{Z} \in \Xi$.*

1. *Let heterogeneity be restricted* ***only*** *to the initial endowments, distributed according pdf $f(\cdot)$ on $[\underline{k}, \overline{k}]$ with mean K_{-1}. Denote k_{-1}^{m} the agent with the median wealth. If $k_{-1}^{m} \in S_{Z,\hat{Z}}$, then either $[\underline{k}, k_{-1}^{m}] \subseteq S_{Z,\hat{Z}}$ or $[k_{-1}^{m}, \overline{k}] \subseteq S_{Z,\hat{Z}}$.*

2. *Let agents be heterogeneous with respect to both labor skills and initial endowment. Also suppose that the initial endowment is an affine function of the skills. If $\theta^m \in S_{Z,\hat{Z}}$ then either $[\underline{\theta}, \theta^m] \subseteq S_{Z,\hat{Z}}$ or $[\theta^m, \overline{\theta}] \subseteq S_{Z,\hat{Z}}$.*

Proof: Due to its simplicity, we present here the proof for the logarithm case. In the appendix 1 we show the proof for the general case.

Take any $Z, \hat{Z} \in \Xi$; we shall show that $J_\varphi(\theta^m) - J_{\hat{\varphi}}(\theta^m) \geq J(\hat{Z}) - J(Z)$ implies $J_\varphi(\theta) - J_{\hat{\varphi}}(\theta) \geq J(\hat{Z}) - J(Z)$ for at least 50% of the agents. A sufficient condition for this to happen is the function $J_\varphi(\theta) - J_{\hat{\varphi}}(\theta)$ being monotone.

We start with the first statement. Given the remark after (12), heterogeneity only with respect to initial endowments allow us to write $\varphi(Z; k_{-1}) = Bk_{-1} + C + D$. Moreover, condition 3) in Lemma 1 implies $\varphi(Z; k_{-1}) = 1 + B(k_{-1} - K_{-1})$.

Given Z and $\hat{Z}$, without loss of generality assume $B > \hat{B}$. If $k^m_{-1} \geq K_{-1}$, then $J_\varphi(k^m_{-1}) \geq J_{\hat{\varphi}}(k^m_{-1})$ and $J_\varphi(k_{-1}) \geq J_{\hat{\varphi}}(k_{-1})$ for all $k_{-1} \geq k^m_{-1}$. If $k^m_{-1} \leq K_{-1}$, then $J_\varphi(k^m_{-1}) \leq J_{\hat{\varphi}}(k^m_{-1})$ and $J_\varphi(k_{-1}) \leq J_{\hat{\varphi}}(k_{-1})$ for all $k_{-1} \leq k^m_{-1}$. This proves the first statement.

Next, consider the case with heterogeneous labor skills: $\varphi(Z, \theta) = Bk_{-1}(\theta) + C\theta + D$. We use the fact that the initial endowments are an affine function of the skill level, $k_{-1}(\theta) = \eta_1 + \eta_2\theta$. Then:

$$\frac{\partial(J_\varphi(\theta) - J_{\hat{\varphi}}(\theta))}{\partial\theta} = \frac{1}{1-\beta}\left[\frac{\text{"constant"} + \Big[[B\eta_2 + C][\hat{B}\hat{\eta}_2 + \hat{C}]\theta\Big] - \Big[[\hat{B}\hat{\eta}_2 + \hat{C}][B\eta_2 + C]\theta\Big]}{[Bk_{-1}(\theta) + C\theta + D][\hat{B}k_{-1}(\theta) + \hat{C}\theta + \hat{D}]}\right] \tag{14}$$

Therefore the sign of the derivative does not depend on θ ∎

There is an obvious abuse of notation in Proposition 1, since the set of implementable allocations are different depending on the type of heterogeneity in the economy.

Next we highlight the key factors behind the proof of Proposition 1. First, as mentioned before, given homothetic preferences, interior individual allocations of consumption and leisure are proportional to the counterpart aggregates. Therefore when comparing allocations Z and $\hat{Z}$, what is key is the ratio of the proportionality factors $\varphi(\theta)/\hat{\varphi}(\theta)$. Moreover, under the full insurance assumption, the proportionality factors are constant over time and are given by the value of the after tax total wealth that individuals would have if they would sell the full amount of labor to the firms. Under the assumption on the affine tax schedule for labor income, the after tax human wealth is linear in the productivity type. If there is no initial wealth inequality the function $\varphi(\theta)/\hat{\varphi}(\theta)$ is monotone in the productivity type, and therefore if the median type θ^m prefers Z to $\hat{Z}$ then at least half of the remaining types will also agree on the ordering over these two allocations. In the case of initial wealth heterogeneity, one way to ensure that the result holds is to assume that initial endowments are an affine function of the skills.

It is important to emphasize that the role of the affine tax schedule assumption is central to the above construction. In particular, it is key the fact that the after tax human wealth is linear in the productivity type. Certainly this would not be true for a general class of nonlinear tax schedules.

Next, we briefly discuss another environments in which the consensus result can be replicated. First, it is challenging to relax the assumption on the homotheticity of preferences. The main reason is that small perturbations on preferences would make the whole distribution of after-tax wealth in the economy to matter significantly.

Proposition 1 also would be true in an environment in which there is no lump-sum component in the fiscal policy, but the government collects the taxes revenues in each period and redistributes it trough a public good g_t. In this case, the utility $v(g_t)$ that individuals get from g_t should enter additively in the period utility function.

Given the linearity restriction imposed in the Proposition 1, we shall assume the following.[7]

Assumption 1: The initial endowments are an affine and increasing function of skills among types: $k_{-1}(\theta) = \gamma_k + (K_{-1} - \gamma_k) \cdot \theta$ with $0 \leq \gamma_k \leq K_{-1}$.

Finally, without a restriction on the value of the lowest type $\underline{\theta}$, Proposition 1 establishes the consensus result only for fiscal policies that support interior equilibria. Without any such restriction, for some fiscal policies there will be aggregate allocations in which the decentralized competitive equilibrium exhibits a positive measure of agents supplying zero labor in equilibrium. Usually such aggregate allocations have the feature that the lump-sum component of the tax schedule is too large (a positive transfer), making too costly for the lowest types to work. These types will better off not working at least in some periods. For more details see Piguillem and Schneider (2007). In order to avoid considering economies in which non-interior allocations exist, we present a lemma that will be used to impose a lower bound on the value of $\underline{\theta}$.[8]

[7]The linearity condition does not mean that the initial distribution of capital is linear itself.

[8]The analysis would be much more complicated in this case.

Lemma 2. *Consider any Z satisfying conditions (1)-(2) in Lemma 1 and having both aggregate labor sequence bounded away from zero and $\widetilde{W}_0(\theta, T, \tau_0)/\theta > 0$. There exists $\widehat{\theta} < 1$ such that $\varphi(Z; \theta) \leq \frac{1}{1-L(s^t)}$ for all s^t, $\theta \geq \widehat{\theta}$, and $L(s^t) \subset Z$.*

Proof: See the appendix 1.

The lemma above provides a minimum value for $\underline{\theta}$ such that, even for the maximum feasible level of transfers $-\overline{T} > 0$ (in a competitive equilibrium with aggregate labor bounded away from zero), the lowest type will work a positive amount in any period and state of nature. It also imposes a restriction on the variance of the distribution of skills.

By assumption 1, equation (11) states that individual labor supply is a monotone function in θ. Notice that the condition $\widetilde{W}_0(\theta, T, \tau_0)/\theta > 0$ implies that individual labor is minimum for the lowest type $\underline{\theta}$ in all periods and states.[9]

Assumption 2: $\underline{\theta} \in [\widehat{\theta}, 1)$.

[9] One may ask what could happen in cases in which $\widetilde{W}_0(\theta, T, \tau_0)/\theta < 0$. We believe that in such cases the labor supply will be strictly positive for the highest type. In this case we can show there exists $\widetilde{\theta} > 1$ such that the upper bound constraint in Lemma 1, part 3, will never bind. Furthermore, for the type of distribution that we analyze in the next section, the median voter will indeed prefer $-T \geq 0$.

5 Characterization of the Condorcet Winner in the log case

The characterization of the Condorcet winner comes from the maximization of the utility for the type θ^m given that the agent has to pick a sequence of aggregate allocations, an initial tax on capital and lump-sum transfers that can be supported as a competitive equilibrium. Lemma 1 gives us the sufficient implementability conditions that should be satisfied. Assumption 1 implies that is sufficient to check only the non-negativity constraint for the lowest type $\underline{\theta}$.

Then we shall partially characterize the solution for the following problem:

$$\mathbf{P(M)}: \quad \max_{\{C,L,K,T,\tau_0\}} \left\{ \frac{1}{1-\beta} \log\left(1 + (1-\beta)\left[\left(\widetilde{W}_0(\theta^m, T, \tau_{k0}) - \mathrm{E}(\widetilde{W}_0(\theta, T, \tau_{k0}))\right)\right.\right.\right.$$
$$\left.\left.\left. + (\theta^m - 1)\cdot UL\right]\right) + \sum_{t,s^t} \beta^t \Pr(s^t)\left[\alpha \log(C(s^t)) + (1-\alpha)\log(1 - L(s^t))\right]\right\}$$

$$\text{s.t.} \begin{cases} C(s^t) + K(s^t) \leq F\left(L(s^t), K(s^{t-1}), s^t\right) + (1-\delta)K(s^{t-1}) & \forall s^t \quad (\mathbf{RC}); \\ \frac{1}{C(s^t)} \geq \beta \sum_{s_{t+1}} \Pr(s_{t+1}|s^t) \frac{[1+(1-\overline{\tau})(F_k(s^{t+1})-\delta)]}{C(s^{t+1})}, & \forall s^t \quad (\mathbf{UB}); \\ 1 + (1-\beta)\left[\left(\widetilde{W}_0(\underline{\theta}, T, \tau_{k0}) - \mathrm{E}(\widetilde{W}_0(\theta, T, \tau_{k0}))\right) + (\underline{\theta} - 1)UL\right] \geq 0 & (\mathbf{NN}); \\ \tau_{k0} \leq \overline{\tau}, \ K_{-1} \text{ given} & \end{cases}$$

If some agent were given the power to choose an implementable allocation, she would care about her own proportion of the aggregate allocations and also about the utility for the mean type. Those are the two parts in the objective function of the (median voter) problem. The "proportional part" basically depends on the difference in after tax total wealth between the mean and the median type. As the proof of Lemma 3 below shows, this share will be higher (lower) than θ^m if the tax schedule includes positive lump-sump transfers (taxes).

Remark: If in the solution to P(M) we have that $T \leq 0$ then, using the fact that $E(\varphi(Z;\theta)) = 1$, constraint (NN) will not bind.

Next we state the first result relating the size of taxes and the distance between the median and the mean type. Basically, if the distance is zero, and if fiscal policy has only redistribution concerns (zero government spending), then taxes are zero in all periods and states.

Proposition 2. *Suppose $\theta^m = E(\theta) = 1$. Under Assumption 1, the most preferred allocation for the median type is the solution to a version of the Neoclassical Growth model in this environment with a representative agent having labor productivity equal to the unity and endowment K_{-1}. The implied taxes are given by $\tau_l(s^t) = \tau_k(s^t) = 0$ for all s^t.*

Proof: First, by assumption we have $E[\widetilde{W}_0(\theta, T, \tau_{k0})] = \widetilde{W}_0(\theta^m, T, \tau_{k0})$. But then the objective function reduces to:

$$\max_{C,L,K} \sum_{t,s^t} \beta^t \Pr(s^t) \left[\alpha \log(C(s^t)) + (1-\alpha)\log(1 - L(s^t))\right]$$

Finally we show that maximizing the above objective function subject to the RC constraint only satisfies all the remaining constraints, i.e., UB, and NN. As it is well known the solution for the above problem implies no taxation. Since there is no initial government debt by assumption, we get that $T = 0$. This immediately implies that constraint (NN) is satisfied. But since $\tau_k(s^t) = \tau_l(s^t) = 0 \;\forall\; s^t$ we have that UB is not binding ∎

The only way that the median voter can take advantage of nonzero marginal taxes is through the difference between the value of his wealth (initial wealth and the market value of labor endowment) and the mean wealth. When such a difference does not exist, marginal taxes are always zero. In the case where there is a sufficiently small process for government spending, the chosen fiscal policy will use only lump-sum taxes to finance the stream of expenditures. We refer to a sufficiently small process because otherwise the poorest individual in the economy may not afford the payment of the lump-sum tax.

Other than the result about taxes, the claim above adds a new interpretation to the Neoclassical Growth Model with homothetic preferences. That is, its solution can be thought also as the aggregate competitive equilibrium allocation that would be chosen by majority voting at time zero in an economy having heterogeneous labor skills drawn from a non-skewed distribution.

Next we turn the case where $\theta^m \neq E(\theta)$. The next result, Lemma 3, will be important later. It states that constraint NN will never bind in the solution to P(M).

Lemma 3. *If $\theta^m < 1$ then $T \leq 0$ in any solution to P(M).*

Proof: See the appendix 1.

The intuition for Lemma 3 is simple. The main objective of the median voter is to achieve some degree of redistribution in her favor. This occurs only when the agent receive more resources than she pays. That is, because all the distortive taxes are linear and since $\theta^m < E(\theta)$, she always pays (receives, if taxes are negative) less than the average agent. On the other hand, given that all agents receive (pay) the same transfer, the only way for her to get some benefit from redistribution is to set the revenues from linear taxation at a positive value (pay less than the average) and the lump sum at a negative value (receiving the same as the average). When there is no government spending the difference is a net gain for the median agent.

Next we state a lemma which will be used later in Proposition 3. The result is an extension of the capital tax result in Bassetto and Benhabib (2006). In the next lemma, the notation $s^t > s^{\tilde{t}}$ is supposed to be understood as the histories that immediately follow s^t.

Lemma 4. ***(The Bang-Bang Property)*** *In the solution for the median voter's problem, if there exists $s^{\tilde{t}}$ such that the implied tax $\tau_k(s^{\tilde{t}}) < \overline{\tau}$ then*

$$\frac{1}{C^*(s^t)} = \beta \sum_{s_{t+1}} Pr(s_{t+1}|s^t) \frac{[1 + F_k^*(s^{t+1}) - \delta]}{C^*(s^{t+1})} \ \ \forall \ s^t > s^{\tilde{t}}$$

and therefore $\tau_k(s^t) = 0$ for all $s^t > s^{\tilde{t}}$.

Proof: See the appendix 1.

Remark: Notice that the proof above depends on the return function in P(M) being increasing in the utility of the mean type. In the general case ($\sigma \neq 1$), this may not be true, as Bassetto and Benhabib (2006) shows. When the return function is decreasing in the utility of the mean type, the proof can be adapted to show that the constraint UB is always binding.

Next we show that, when $\delta = 0$ and heterogeneity in the initial distribution of capital is sufficiently small, taxes on capital will be zero for any period $t \geq 2$.

Lemma 5. ***(Capital taxes)*** *Suppose $\theta^m < E(\theta) = 1$ and $\delta = 0$. Under assumption 1, there exists $\epsilon > 0$ such that for all K_{-1} and γ_k with $|K_{-1} - \gamma_k| \leq \epsilon$, the implied capital taxes in the solution to P(M) are given by:*

$$\tau_k(s^t) = \begin{cases} \overline{\tau} & \textit{if } t = 0 \\ 0 < \tau_k(s^t) \leq \overline{\tau} & \textit{if } t = 1 \\ 0 & \textit{if } t \geq 2 \end{cases}$$

Proof: See the appendix 1.

If the value of $K_{-1} - \gamma_k$ is not small enough, then the Lemma must be modified slightly. Instead of having $\tau_k(s^t) = 0$ for all $t \geq 2$ it would be true for all $s^t > s^{\tilde{t}}$ given some finite $\tilde{t}$. This is a very well known result dating from the original work of Chamley (1986). It follows from the fact that, otherwise, the solution would exhibit $U^*_{ct} = \beta E_t[U^*_{c_{t+1}}]$ $\forall t$. Since any solution should have $U^*_{ct}(s^t) < \infty$ $\forall$ s^t, and therefore $E(U^*_{ct}) < \infty$ for all t, it follows by the law of iterated expectations that $U^*_{ct} = \lim_{T\to\infty} \beta^T E_t[U^*_{c_{t+T}}]$. Since it can be shown that $\{U^*_{ct}\}_t$ is a submartingale, we can use Dobb's convergence theorem to show that this limit exists and is equal to zero. This leads to a contradiction since the constraint set is compact in the product topology. As we have pointed out in the remark right after Lemma 4, for the general Cobb-Douglas utility function the constraint UB may bind always.

Now consider the labor income tax. From the competitive equilibrium we know $F_L(s^t)(1-\tau_l(s^t)) = \frac{1-\alpha}{E(\theta)-L(s^t)}\frac{C(s^t)}{\alpha}$. Therefore $1-\tau_l(s^t) = \frac{(1-\alpha)}{E(\theta)-L(s^t)}\frac{C(s^t)}{F_L(s^t)\alpha}$. Then we have the following.

Proposition 3. ***(Labor Tax)*** *Suppose that $\theta^m < 1$. Then in the solution to P(M) there exists a history $s^{\hat{t}}$ such that, for all $s^t > s^{\hat{t}}$ the implied labor taxes are:*

1. $0 < \tau_l(s^t) < 1$.

2. *$\tau_l(s^t)$ depends on s^t only through $L(s^t)$:*

$$\tau_l(s^t) = \frac{(1-\theta^m)}{\varphi(Z^*;\theta^m)(1-L^*(s^t)) + (1-\theta^m)}$$

3. $\frac{\partial \tau_l(s^t)}{\partial(1-\theta^m)} > 0$.

Proof: The existence of a history $s^{\hat{t}}$ in which UB stops binding was justified in the previous paragraph.

Recall from Lemma 3 that (NN) is not binding. Let $\lambda(s^t)$ be the lagrange multipliers associated with (RC). Then the first order condition with respect to aggregate labor is:

$$\left[\frac{(1-\alpha)(\theta^m - E(\theta))\beta^t \Pr(s^t)}{\varphi(\theta^m)[1-L(s^t)]^2}\right] - \frac{(1-\alpha)\beta^t \Pr(s^t)}{[1-L(s^t)]} + \lambda(s^t) F_L(s^t) = 0 \text{ for } t \geq 1 \quad (15)$$

The implied tax on labor is given by:

$$1 - \tau_l(s^t) = \left[\left(\frac{(1-\theta^m)}{\varphi(\theta^m)}\right)\frac{1}{1-L(s^t)} + 1\right]^{-1} \quad (16)$$

Let $H = \frac{(1-\theta^m)}{\varphi(\theta^m)} > 0$. Then (16) can be rewritten as:

$$0 < \tau_l(s^t) = \frac{H}{1-L(s^t)+H} < 1 \text{ for } t \geq 1$$

or

$$\tau_l(s^t) = \frac{1}{\left[\frac{1}{1-\theta^m} - (1-\beta)\left(\frac{\alpha R_0(K_{-1}-\gamma_k)}{C_0} + UL\right)\right][1-L(s^t)] + 1}$$

It is straightforward to check that $\frac{\partial \tau_l(s^t)}{\partial(1-\theta^m)} > 0$ ∎

Corollary 1. *(**Extending Bassetto and Benhabib (2006)**) Suppose that heterogeneity is restricted only to the initial wealth distribution, that is, agents are indexed by the parameter $k \in [\underline{k}, \overline{k}]$ distributed with p.d.f. $f(\cdot)$ and the labor skill is given by $\theta(k) = 1 \; \forall \; k \in [\underline{k}, \overline{k}]$. Then there exists $s^{\hat{t}}$ such that the Condorcet winner has $\tau_l(s^t) = 0$ for all $s^t \geq s^{\hat{t}}$.*

Both Lemma 5 and Proposition 3 provide some explanation of how heterogeneity shapes the Condorcet winner when agents have log preferences. First, as the median voter's problem illustrates, her payoff depends positively on the payoff obtained by the mean type. This implies that nonzero taxes benefits the voter only to the extend to which she can manipulate her share $\varphi(\theta^m)$ through the lump-sum transfers. The mean type always prefers zero taxes, and the share $\varphi(\theta^m)$ depends on the difference in after tax wealth between the median and the mean type.

Since positive capital taxes reduce the payoff for the mean type, Lemma 5 implies that the first two periods in the economy are sufficient to obtain full benefits from positive taxes on capital if the initial heterogeneity of capital endowments is small. On the other hand, labor taxes are always positive since the heterogeneity in labor income never disappears.

Also we find a smoothing effect on labor taxes similar to Werning (2007). Since distortions decrease the utility for the mean type, concavity implies that labor taxes should be higher in states in which aggregate labor is higher.

In the appendix 1, the results about taxes in the general case ($\sigma \neq 1$) are extended. As Bassetto and Benhabib (2006) point out, depending on the magnitude of σ, capital taxes may be always at the upper bound.[10]

Taking this into consideration, we find a condition on the size of σ such that the capital taxes eventually go zero. This helps to characterize taxes on labor.

The results presented in the appendix 1 are very close to those in the last section with slight modifications. Instead of propositions for every $t \geq 2$, we have results

[10] The proof is omitted because it is just an extension of the reasoning presented in Bassetto and Benhabib (2006).

for all $t > \hat{t}$, for some $\hat{t} \geq 2$. In addition, the statements are weaker in the sense that they depend on σ being smaller than $(1-\theta^m)^{-1}$. The main role of the condition is to make sure that the objective function is increasing in aggregate consumption and decreasing in aggregate labor, as Lemma 9 shows.

The results are summarized as follows. Provided that the inequality in skills is not too large or, alternatively, $\sigma \leq (1-\theta^m)^{-1}$, capital income taxes will eventually be zero. Labor income taxes are always positive, increasing in inequality and state dependent. For all histories after which the upper bound constraint on capital taxes is not binding, labor taxes are given by:

$$\tau_l(s^t) = \frac{(1-\theta^m)}{(\theta^m - 1)[(1-\alpha)(1-\sigma)-1] + [1-L(s^t)][(1-\sigma)+\sigma\chi(Z,\theta^m)]} \tag{17}$$

where $\chi(Z,\theta^m)$ resembles the proportionality factor $\varphi(Z;\theta)$ in the log case.

6 Stochastic Labor Skills

In this section we extend the previous environment to an economy where types are fixed but labor skills evolve stochastically over time. Moreover, the extension is done in such a way that we can apply the previous consensus result. As before, types are initially distributed according to a skewed distribution on $\Theta = [\underline{\theta}, \overline{\theta}]$. Each type is related to an initial skill θ_0^i. The only modification in the physical environment is the following. For each period after t=0, skills of type i evolve stochastically, and are potentially correlated with the aggregate state. For each history s^t, skills are given by: $\theta^i(s^t) = \gamma(s^t) + \rho(s^t)\theta_0^i$. This specification allows correlation between changes in the distribution of skills and aggregate productivity shocks. In addition, $\rho(s^t)$ and $\gamma(s^t)$ may be chosen such that the changes are a mean preserving spread of any particular state s_t.

As before, individual allocations are proportional to aggregate allocations. In this economy the individual shares are given by:

$$\varphi(Z;\theta_0^i) = 1+(1-\beta)\left[\left(\widetilde{W}_0(\theta_0^i, T, \tau_{k0}) - \mathrm{E}(\widetilde{W}_0(\theta_0^i, T, \tau_{k0}))\right) + (\theta_0^i - 1)\sum_{s^t} \rho(s^t)U_L(s^t)\right] \tag{18}$$

where $U_L(s^t) \equiv \frac{(1-\alpha)\beta^t Pr(s^t)}{E_{s_t}(\theta) - L(s^t)}$.

At this point it should be clear that the consensus result also holds in this economy because the linearity restriction on the initial types remains.

Next we use a simpler example to highlight the effects of this specific stochastic skill process on labor taxes chosen by the median voter. In an economy without capital accumulation, the best marginal tax on labor income for the median type is given by:

$$\tau_l(s^t) = \frac{(1-\theta_0^m)\rho(s^t)}{\varphi(\theta_0^m)(1-L(s^t)) + (1-\theta_0^m)\rho(s^t)} \tag{19}$$

The larger the distance between the median and the mean type, the higher the labor tax on that state for a given aggregate labor quantity. The final effect on taxes is ambiguous. As equation (19) shows, the result depends on two factors. First, there is a tax smoothing effect: the larger the aggregate labor allocation, other things constant, the higher the tax. This is closely related to concavity and the fact that the median's utility depend on the utility of the mean type. The other effect is related to how the skills' distribution changes over the business cycles (its correlation with aggregate shocks). An increase in the distance between the mean and the median agent increases the gains of redistributive policies for the median voter, and therefore call for higher taxes.

Thus, if inequality and employment are positively correlated, the effects reinforce each other and labor taxes are unambiguously higher. However, if inequality rises in periods of low employment (inequality and employment are negatively correlated) both effects act in opposite directions, turning the sign of the correlation between employment and labor taxation ambiguous. As we pointed out before, the final outcome on taxes depends on two effects, which are illustrated by the numerical example in the next section.

6.1 A Numerical Exercise

Consider an economy without capital accumulation and where skills evolve stochastically as in the previous section. Assume that there are only two possible states in the economy, $S = (High, Low)$. The stochastic process for the states is i.i.d (allowing for persistence will not affect the qualitative results), with $\pi_H = 0.6$ and $\pi_L = 0.4$. The initial state is $s_0 = H$

Technology is linear, $Y(s^t, L(s^t)) = A(s^t)L(s^t)$, with the aggregate productivity parameter being $A_L = 1.25$ and $A_H = 0.95$. Government consumption in each state takes on the values $G_H = G_L = 0.08$, which makes government consumption being about 17% of output. Preferences are logarithmic ($\sigma = 1$). We also set $\alpha = 0.3$ and $\beta = 0.95$.

The initial distribution is skewed, with the mean normalized to one and $\theta^m = 0.9$. In addition, we assume that in the high state the distribution of skills is always the same as the distribution at the initial period ($\rho(H) = 1$ and $\gamma(H) = 0$). In the low state, the distribution of skills is a mean preserving spread of the distribution at the initial state ($\rho(L) > 1$ and $\gamma(L) < 0$ with $\rho(L) + \gamma(L) = 1$).

In order to analyze the impact of the changes on the distribution of skills on the cyclical properties of the fiscal policy, we consider two economies: in Economy 1 the distribution of skills has low variability. The economy parameterized by $\rho_1(L) = 1.02$ and $\gamma_1(L) = -0.02$. In Economy 2, the skills distribution is more volatile than Economy 1. We achieve this by setting $\rho_2(L) = 1.05$ and $\gamma_2(L) = -0.05$. Figure 1 shows for the two economies the distribution at time zero and the distributions when aggregate state is $s = low$.

Next we show the calculated taxes for each state.

State	Economy I		Economy II	
	τ	L	τ	L
High	0.279	0.30	0.28	0.30
Low	0.278	0.285	0.285	0.28

Table 1: Aggregate labor and marginal labor tax in each state.

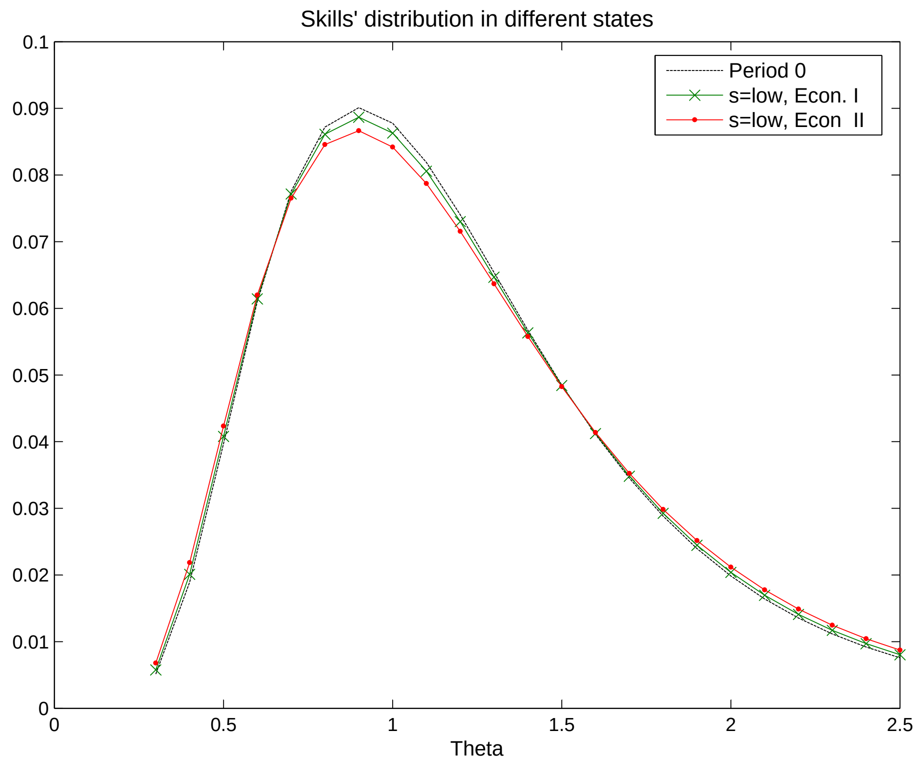

Figure 1: Skill distribution in period zero and in state $s = low$.

Labor taxes will be either procyclical or counter-cyclical, depending on how the distribution of heterogeneity changes in the low state. In Economy I, the smoothing effect predominates: higher aggregate labor implies higher taxes. In Economy II, the larger distance between the median and the mean type causes the labor tax be higher for states in which aggregate labor is lower.

The example illustrates the potential ability of the model to explain differences in the business cycles properties of labor taxes. If rich countries have sufficiently smaller dispersion in the skill distribution during bad times than poor ones, then rich countries will exhibit counter-cyclical labor taxes, while poor countries will have pro-cyclical labor taxes.

7 Quantitative Results

Several papers like Eckstein and Nagypal (2004), Heathcote, Storesletten, and Violante (2008), among others, have reported the increasing trend in labor income inequality in the U.S. in the last decades. Regarding labor income taxation, McDaniel (2007) constructs average taxes for the U.S. (and OECD countries) for the period 1950-2003 and it finds an increasing trend.

In this section we show two quantitative results of the model. First, through equation (17), we calculate both a lower and an upper bound on how much of the increase in labor taxes observed in the data can be accounted by the model. Second, we numerically solve the median voter's problem using a simple calibrated version of the model with stochastic labor skills. Then we compare the correlation between labor taxes and labor allocations (and GDP growth) from the model with the data.

Since the model does not have consumption taxes, we follow the same methodology as in Ohanian, Raffo and Rogerson (2006). We compare the labor taxes from the numerical solution with $1 - \frac{1-\tau_{lt}}{1+\tau_{ct}}$ from the data. Figure 2 shows the trends in the data.

Figure 2: Average labor taxes and ratio mean to median earnings in the US (calculated using data in Eckstein and Nagypal (2004)).

7.1 Data and Calibration

We take the average taxes on both labor and consumption for the US economy in the period 1950-2003 from McDaniel (2007).

Uncertainty is described by a Markov chain with 4 states. Both TFP and labor skills shocks are assumed to take two possible values: $A(s) \in \{A_L, A_H\}$ and $\rho(s) \in \{\rho_L, \rho_H\}$. The data for the macroeconomic aggregates are from NIPA, in billions of chained 2000 dollars covering the period 1960-2006. Since in the model we normalize the endowment of time to be equal to one, we construct a new labor series as the ratio between the total average weekly hours worked from BEA and the potential number of hours (5200 times population of 16 and over).

The production function is Cobb-Douglas with capital share $\nu = 0.3$, a usual value found in the literature. The technology parameter $A(s)$ is calibrated by using GDP from NIPA and labor as the total average weekly hours worked. The skill distribution parameters ρ_H and ρ_L are calculated from the wage inequality data in Eckstein and Nagypal (2004).[11] We take mean and median wages as a proxy for mean and median individual skills respectively. One drawback is that the data refers to weekly earnings, and therefore it does not account for the effects of cross section variation of hours worked. But since we only need data about mean and median wages, and also given that aggregate hours worked in US has been quite stable over the last decades, we think this issue is not very critical for our purposes.

We calibrate the transition matrix over the possible four states by filtering both the TFP and the ratio median to mean wages. Then we calculate the probabilities using the frequency of the states observed in the data. The matrix below show the calculated probabilities over S=$\{s1 = (A_H, \rho_H), s2 = (A_H, \rho_L), s3 = (A_L, \rho_H), s4 = (A_L, \rho_L)\}$.

[11] The authors use data from the Current Population Survey covering the period between 1961 and 2002.

$$\begin{pmatrix} 0.438 & 0.125 & 0.375 & 0.062 \\ 0.286 & 0.286 & 0.142 & 0.286 \\ 0.455 & 0.0 & 0.091 & 0.454 \\ 0.20 & 0.20 & 0.30 & 0.30 \end{pmatrix}$$

Transition matrix.

We consider $A_H = 1 + \varepsilon_H$ and $A_L = 1 - \varepsilon_L$. We choose ε_H and ε_L such that the unconditional mean is equal to one and the process matches the variance of GDP growth in the data. In this way we set $\varepsilon_H = 0.004$ and $\varepsilon_L = 0.0064$.

The depreciation rate is set to the usual value of 0.06. We use log preferences. The parameter α is set to match, on average during the period considered, the first order conditions in the median voter's problem. Using this criterion, we find $\alpha = 0.38$.[12]

Taking into account the average ratio median skill to mean skill equal to 0.79, we calibrate ρ_H and ρ_L such that we have the unconditional mean equal to one and the variance of the ratio median to mean wages matches the data. The values for ρ_H, ρ_L and the remaining parameters are summarized in Table 2 below.

We do not consider initial heterogeneity in wealth. Such strong restriction on

[12]Since the first order conditions contain the share of the median type coming from the solution of his problem, the calibration for α is done in two steps. In the first step, we guess a value for the share and then calculate the average α. In the last step, using the calculate value for α, we check if the resulting share is the one that was guessed in the previous step.

Parameter	Description	Value
(α, σ)	preference parameters	(0.38,1)
β	intertemporal discounting	0.96
ν	capital share	0.3
δ	depreciation	0.06
(A_H, A_L)	TFP shocks	(1.004, 0.9936)
θ^m	median skill	0.79
(ρ_H, ρ_L)	skill parameters	(1.054, 0.945)

Table 2: Summary of the calibrated parameters.

the initial wealth distribution is imposed in order to avoid additional complications related to the inequality constraints in the Euler equations. From the theory we know that if $\tau_k(s^1) < \overline{\tau}_k$, then $\tau_k(s^t) = 0$ for all $t > 1$. Since our main concern is about labor taxes, we think that initial wealth heterogeneity would add little content to the discussion at a large cost in terms of computational issues. Since the problem is not recursive, we solve it using a two-step algorithm that explores the recursive property of the Lagrangean. For more details see section A7 in the appendix 1.

The main finding of the calibrated model is a good fit of the trend in labor taxes. We assume that the conditions of Proposition 3 holds, so that labor taxes are given by (using the extension of the stochastic labor skills case):

$$\tau_l(s^t) = \frac{\rho(s^t)(1-\theta_0^m)}{\varphi(Z^*;\theta^m)(1-L(s^t)))+\rho(s^t)(1-\theta_0^m)}$$

Assuming that $\rho_t(1-\theta_0^m)$ is the actual realization of $\rho(s^t)(1-\theta_0^m)$, we can calculate both an upper and a lower bound on the process for labor taxes. These bounds come from the proof of Lemma 3 in the appendix 1: in the solution to the median voter's problem, his share is less than the unit and larger than the initial realization in skills. In order to minimize the effects of the choice of the initial period, we set $0.79 \leq \varphi(Z^*;\theta^m) < 1$, where the lower bound is given by the average value of skills in the data.

In Figure 3 we show the bounds on the process for labor taxes. Since we do not use the numerical solution of the model to calculate these boundaries, we have chosen to set the aggregate labor allocation equal to its values calculated in the data. If instead we set the aggregate labor allocation to be equal to the average value in the data, the picture would be very similar.

	Growth (%)			
	Data		Upper	Lower
	Original Taxes	Modified Taxes	Bound	Bound
1962-2001	66	33	63	68
1962-2003	43	20	67	72

Table 3: Increase in labor taxes accounted by the model.

If we consider the period 1962-2001[13], the model accounts for about two times

[13]This specific period does not include the significant decrease in average labor taxes between

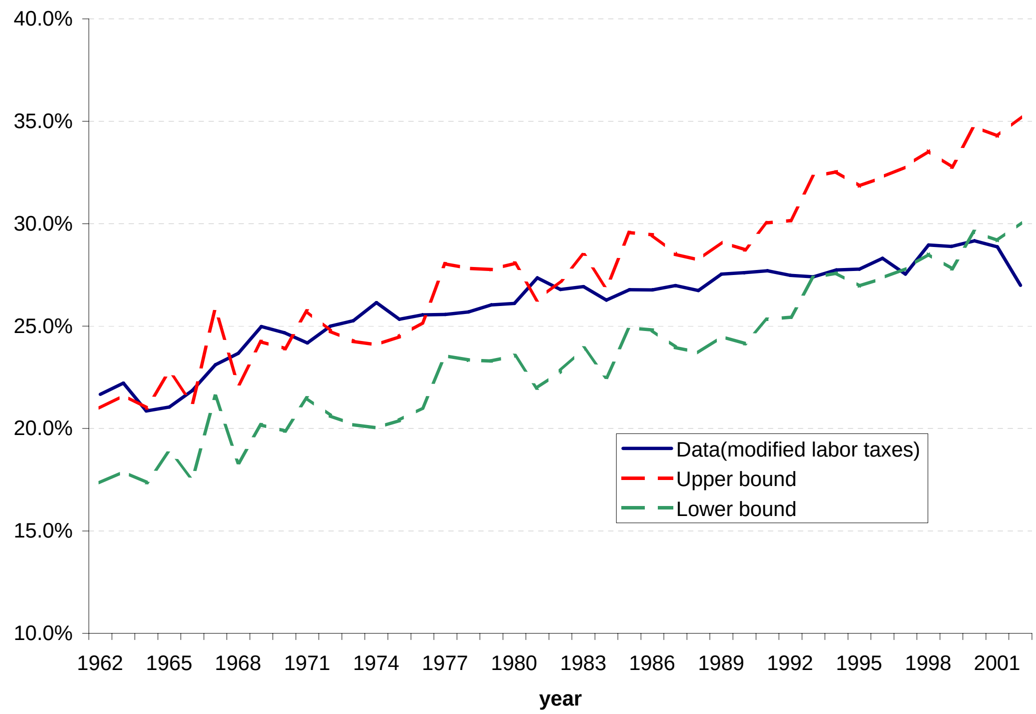

Figure 3: Bounds on average labor taxes in the calibrated model.

the growth of labor taxes observed in the data. In appendix 3 we show the same picture, but for the case where $\sigma = 2$.

Next we highlight some statistical properties of the calibrated economy. Moreover, we compared some properties in the data with two specifications of the calibrated model. In the first specification, labor skills are constant over time. The second specification is the one with stochastic labor skills. As one can see in Table 3 below, the model with constant skills yields almost zero variation in labor taxes, in line with the findings in Chari, Christiano, and Kehoe (1994). When we feed in

2002 and 2003.

the model the variation in skills to match the variance of the ratio median to mean wages, the economy matches the signal of the comovements between taxes and the relevant aggregates. The discussion in section 6 points out that the model without skills shocks should yield a correlation between labor and taxes equal to one. Since in the data the correlation between inequality of wages and TFP is negative, a priori the sign of the correlation between labor and taxes is ambiguous. The net effect in the calibrated model changes the correlation between labor and taxes to negative.

	Statistics				Correlations		
	St.Dev. ΔY	St.Dev. L	Mean Tax	St.Dev. tax	τ-ΔY	τ-ΔL	τ-L
Data	**0.02**	**0.01**	**0.26**	**0.02**	**-0.4**	**-0.1**	**- 0.7**
Model Fixed Types	**0.02**	0.002	0.26	0.001	0.7	0.2	1
Model Changing Types	**0.02**	0.006	0.26	**0.009**	-0.6	-0.2	-1

τ : Labor income tax
Y: Aggregate output
L: Aggregate labor

Table 4: Selected statistical properties of the model, $\sigma = 1$.

Finally, it turns out that the signal of the correlations presented in Table 4 are sensitive to the chosen value for σ. Table 5 shows the correlations when $\sigma = 2$ and the appropriate change in the values for (ρ_H, ρ_L) is done.

	Statistics				Correlations		
	St.Dev. ΔY	St.Dev. L	Mean Tax	St.Dev. tax	τ-ΔY	τ-ΔL	τ-L
Data	**0.02**	**0.01**	**0.26**	**0.02**	**-0.4**	**-0.1**	**- 0.7**
Model Changing Types	**0.02**	0.005	0.26	**0.01**	0.7	0.2	0.9

τ : Labor income tax
Y: Aggregate output
L: Aggregate labor

Table 5: Selected statistical properties of the model, $\sigma = 2$.

8 Conclusion

In this paper we show how heterogeneity shapes redistributive fiscal policy when individuals have balanced growth preferences and are heterogeneous with respect to both labor skills and initial wealth.

We show that the best tax sequence for the type with the median labor productivity cannot be defeated by any other policy. If only one dimensional heterogeneity is considered, i.e., either labor productivity or initial capital heterogeneity, no additional assumption regarding the distribution of types is needed. When both types of heterogeneity are taken into account simultaneously, a linear restriction about the initial wealth is required.

Regarding the characterization of the most preferred allocation by the median type, we show that if her skill is less than the mean, labor taxes are state dependent and always positive. Using a partial derivative argument at the solution, we show that labor taxes are increasing in the distance between the mean and median labor productivity. The results regarding the capital taxes are the same as in Bassetto and Benhabib (2006): taxes are always either zero or at the upper bound.

Through most of the paper we assume that skills are constant, which implies that inequality is independent of the economy's aggregate state. When skills evolve stochastically over time, but preserve the ranking among agents, a temporary increase in inequality could imply either higher or lower labor taxes, depending on both the sign and level of the correlation between inequality and aggregate labor. In an economy without capital accumulation, we present a numerical example where both cases can occur. In the calibrated exercise, we find that the model matches both the increasing trend and the levels of labor taxes observed in US in the last decades. The model accounts for twice as much of the growth in labor taxes observed in the period 1962-2001. The logarithm specification of the model also matches the negative signal of the comovements between taxes and labor and output. We found that the values of such comovements are very sensitive to the value of σ. For values of σ close to 2, the signal of the comovements are positive.

The findings presented here may be useful for economies in which voting occurs sequentially over time. Also the strategy of the proof for the median voter result may be used in economies in which agents decide over objects other than taxes and preferences are homothetic.

Part III

A Median Voter Theorem without Commitment[14]

9 A Generalization of The Consensus Result

In this section we generalize the consensus result presented in Chapter 2 when society does not have a commitment device. Formally, we consider the equilibrium definition in Bernheim and Slavov (2008). For the sake of exposition, we consider the log utility case, full depreciation of capital and no uncertainty. The only restriction that we assume is that the economy lasts for a finite number of periods T. This is indeed important because the main result in proposition 4 currently relies on the interiority of competitive equilibria, and eventually it becomes non-interior in the off-path of the play for arbitrarily distant periods in the future.[15]

It is also convenient to slightly change the notation regarding individual allocations. In this way, consider the continuum of agents in some interval I$\subseteq \mathbb{R}$, with each agent being indexed by the one-dimensional variable $\theta^i \in \Theta \equiv [\underline{\theta}, \overline{\theta}]$. As before, the distribution of types θ^i is represented by the p.d.f. f$(\cdot)$ and the median type is denoted by θ^m. By the linearity assumption used in proposition 1, θ^i represents

[14] This chapter is coauthored with **Facundo Piguillem.**

[15] Details are available upon request.

both her labor skill and a parameter determining the initial capital endowment: $k_0^i = K_0 + (\mathcal{S}_0 - K_0)\theta^i$ for some parameter $\mathcal{S}_0 \in \mathbb{R}$.

We consider sequential markets where the sequence of budget constraints is given by:

$$c_t^i + a_{t+1}^i \leq r_t(1-\tau_{kt})a_t^i + w_t(1-\tau_{lt})l_t^i + T_t$$

where $T_t = F_{kt}K_t\tau_{kt} + F_{lt}L_t\tau_{lt}$ and $a_t^i \in A^i \equiv [\underline{a}, \overline{a}]$ is the bond holdings of agent i. The bounds $[\underline{a}, \overline{a}]$ are chosen to be large enough such that the constraint is never binding.

We write $\Delta \mathbf{a}_t \equiv (\mathbf{a}_t^i)|_{i\in I}$ for the distribution of assets in the economy at the beginning of time t.

10 Dynamic Condorcet Winner Equilibria

Let $h^t = (\tau_0, \tau_1, ..., \tau_t)$ be the histories of previous policies, with $\tau_t \in [-\underline{\tau}_l, \overline{\tau}_l) \times [0, \overline{\tau}_k]$. Let $h_{-1} = \emptyset$ and $H^t = ([-\underline{\tau}_l, \overline{\tau}_l) \times [0, \overline{\tau}_k])^{t+1}$.

Definition 2. *A **fiscal policy(FP) program** is $\pi = \{\pi_t\}_{t\geq 0}$ with $\pi_t : H^{t-1} \to [-\underline{\tau}_l, \overline{\tau}_l] \times [0, \overline{\tau}_k]$ for all $t \geq 0$.*

Let Π be the set of **FP programs.**

Definition 3. *An **allocation mapping** is $\mathcal{A} \equiv (\mathcal{A}^i)|_{i\in I}$ such that $\mathcal{A}^i = \{\mathcal{A}_t^i\}_{t\geq 0}$*

with $\mathcal{A}_t^i : H^t \to \mathbb{R}^3$ *for all* $t \geq 0$.

For each h^t, $\mathcal{A}_t^i(h^t) = (c_t^i(h^t), l_t^i(h^t), a_{t+1}^i(h^t))$.

A pair $(\pi, \mathcal{A})$ generates a unique outcome path $\{\tau_t, \mathcal{A}_t\}_{t\geq 0}$ from the history $h_{-1} = \emptyset$: $\tau_0 = \tau_0(h_{-1})$, $(c_0^i, l_0^i, a_1^i) = \mathcal{A}_t^i(\tau_0(h_{-1}))$ and futures fiscal policy and allocations are generated inductively by $h^t = (h^{t-1}, \pi_t(h^{t-1}))$.

A FP program π and an allocation mapping $\mathcal{A}$ induces after history $h^{t-1} \in H^{t-1}$ a continuation $(\pi, \mathcal{A})|_{h^{t-1}}$. For all $s \geq 0$, $h^s \in H_s$:

$$\pi|_{h^{t-1}}(h^s) = \pi_{t-1+s}(h^{t-1}, h^s)$$

$$\mathcal{A}|_{h^{t-1}}(h^s, \tau) = \mathcal{A}_{t-1+s}(h^{t-1}, h^s, \tau)$$

Given $k_0^i|_{i\in I}$ and a pair $\sigma = (\pi, \mathcal{A})$, we say that the policies and allocations generated by $(\pi|_{h^{t-1}}, \mathcal{A}|_{h^{t-1}})$ is a **competitive equilibrium at time zero** if there exists a sequence of rental prices $\{r_t, w_t\}_{t=0}^{\infty}$ and transfers $\{T_t\}_{t=0}^{\infty}$ such that:

1. Given prices, taxes and transfers generated by $\pi|_{h^{t-1}}$, the sequence of allocations generated by $(\pi|_{h^{t-1}}, \mathcal{A}|_{h^{t-1}})$ solves each agent's problem.

2. Rental prices equal to the marginal outputs.

3. Aggregate resource constraint holds.

Given $a_1^i(h_0)_{i\in I}$ from a CE starting at time zero, a CE starting at history h^t can be defined recursively. For each h^t, there exists an associated distribution of wealth

$\Delta\mathbf{a}_t(h^t)$.

Finally, given a pair $\sigma = (\pi, \mathcal{A})$, a history h^{t-1} and associated distribution of wealth $\Delta\mathbf{a}_t$, let $U^i(\Delta\mathbf{a}_t, \sigma|_{h^{t-1}})$ be the utility attained by household type i under the allocation generated by $(\pi|_{h^{t-1}}, \mathcal{A}|_{h^{t-1}})$:

$$U^i(\Delta\mathbf{a}_t, \sigma|_{h^{t-1}}) \equiv \sum_{s=t}^{T} \beta^{s-t} u^i(c^i_s, 1 - l^i_s/\theta^i) \tag{20}$$

In what follows next, let $\mu(\cdot)$ be the measure over labor skills types.

Definition 4. *Given a pair $(\pi, \mathcal{A})$ and an alternative FF program $\widehat{\pi}$, if in history h^{t-1}*

$$\mu\left\{ i \middle| U^i(\Delta\mathbf{a}_t, \sigma|_{h^{t-1}}) \geq U^i(\Delta\mathbf{a}_t, \widehat{\pi}|_{h^{t-1}}, \mathcal{A}|_{h^{t-1}}) \right\} \geq 1/2$$

*we say that the continuation $\pi|_{h^t}$ **defeats by majority** the continuation $\widehat{\pi}|_{h^t}$ given $\mathcal{A}|_{h^{t-1}}$ and we express $(\pi|_{h^{t-1}}, \mathcal{A}|_{h^{t-1}}) \succeq_M (\widehat{\pi}|_{h^{t-1}}, \mathcal{A}|_{h^{t-1}})$.*

Definition 5. *Let π be a FP program and $\mathcal{A}$ an allocation mapping. We say that $\sigma = (\pi, \mathcal{A})$ is a dynamic Condorcet winner equilibrium (**DCWE**) if:*

(i) $\forall$ $t, h^{t-1} \in H_{t-1}$ with respective $\Delta\mathbf{a}_t$, and for all $\hat{\tau}_0 \in [-\underline{\tau}_l, \overline{\tau}_l] \times [0, \overline{\tau}_k]$, let $(\hat{\tau}, \hat{q})$ be generated by $\sigma|_{(h^{t-1}, \hat{\tau}_0)}$. Then $(\hat{\tau}, \hat{q})$ is a competitive equilibrium.

(ii) $\forall$ $t, h^{t-1} \in H_{t-1}$ with respective $\Delta\mathbf{a}_t$, and $\forall$ $\widehat{\pi}|_{h^{t-1}} \in \Pi$:

$$(\pi|_{h^{t-1}}, \mathcal{A}|_{h^{t-1}}) \succeq_M (\widehat{\pi}|_{h^{t-1}}, \mathcal{A}|_{h^{t-1}})$$

11 Competitive Equilibria Characterization

In what follows below, u^i_c denotes the marginal utility of consumption and $F_k(F_l)$ the marginal product of capital (labor).

Lemma 6. *A sequence* $\{(c^i_t, a^i_t, l^i_t)_{i\in I}\}^T_{t\geq 0}$ *with interior labor and consumption is a competitive equilibrium allocation if and only if:*

1. *evaluated at the aggregate allocations,* $C_t + K_{t+1} \leq F(K_t, L_t)$ $\forall t$;

2. $(1-\bar{\tau}_k)\beta F_{k_{t+1}} u^i_{c_{t+1}} \leq u^i_{ct} \leq \beta F_{k_{t+1}} u^i_{c_{t+1}}$ $\forall t$;

3. $\sum^T_{t\geq 0} u^i_{ct} c^i_t = u^i_{c0}(1-\tau_{k0})F_{k0}k^i_0 + \sum^T_{t\geq 0} u^i_{ct}[(1-\tau_{lt})F_{lt}l^i_t + T_t]$;

4. $\frac{1-\alpha}{\alpha}\frac{L_t}{1-L_t}\frac{C_t}{(1-\gamma)F(K_t,L_t)} \in [1-\overline{\tau}_l, 1-\underline{\tau}_l]$ $\forall t$.

where $T_t = \tau_{kt}F_{kt}K_t + \tau_{lt}F_{lt}L_t$

Proof: (Only If) Let $\widetilde{a}^i_{t+1} \equiv a^i_{t+1} - \underline{a}$. The necessary conditions for the maximization of the individual's problem are :

$$u^i_{ct} = \lambda^i_t,\ u^i_{lt} = \lambda^i_t p_{lt},\ (\lambda^i_t - \lambda^i_{t+1}(1-\tau_{k_{t+1}})r_{t+1})\widetilde{a}^i_{t+1}0$$

and the transversality condition $a^i_{T+1} = 0$.

Using the both the transversality condition and the first order conditions we get (3) for each i. The other conditions are trivial. Notice that the first order conditions can be manipulated such that individual allocations of labor and consumption can be expressed as shares of the aggregates. Moreover, $(1-\tau_{lt}) = \frac{1-\alpha}{\alpha}\frac{C_t}{1-L_t}\frac{1}{F^{\xi}_l}$.

(If) Set $(1-\tau_{lt}) = \frac{1-\alpha}{\alpha}\frac{C_t}{1-L_t}\frac{1}{F_l^{\xi}}$, $r_t = F_{kt}$. Rental prices for labor and bond prices are given by $w_t = F_{lt}$.

Taxes on capital are given by $1 - \frac{u^i_{ct}}{u^i_{c_{t+1}} r_{t+1}}$. By setting $\lambda^i_t = u^i_{ct}$ the first order conditions are met.Construct total wealth levels such that the budget constraint is satisfied with equality in each period:

$$\lambda^i_t(r_t(1-\tau_{kt})a^i_t) = \sum_{s=t}^{T} \lambda^i_s[(1-\tau_{ls})w_s l^i_s + T_s - c^i_s]$$

The resource constraint is met by integrating the budget (which holds with equality) over types in each period ∎

Next lemma shows a similar characterization of the individual shares in lemma 1.

Lemma 7. *The necessary and sufficient conditions (1)-(4) in lemma 6 are equivalent to:*

There exists a function $\varphi^i : \Theta \to \mathbb{R}_+$ *such that:*

1' $C_t + K_{t+1} = F(K_t, L_t)$ *with* $[c^i_t, \theta^i - l^i_t] = \varphi^i[C_t, 1-L_t]$

2' $(1-\bar{\tau}_k)\beta F_{k_{t+1}} u_{c_{t+1}} \le u_{ct} \le \beta F_{k_{t+1}} u_{c_{t+1}}$

3' $\varphi^i = \frac{(1-\beta)}{1-\beta^{T+1}}\left[\frac{\alpha}{C_0}(1-\tau_{k0})F_{k0}k^i_0 + \sum_{t\ge 0}^{T}\frac{\alpha\beta^t}{C_t}((1-\tau_{lt})F_{lt}\cdot\theta^i + T_t)\right]$ *and* $\sum_i \mu_i \varphi^i = 1$

4' $\frac{1-\alpha}{\alpha}\frac{L_t}{1-L_t}\frac{C_t}{(1-\gamma)F^{\xi}_{lt}} \in [1-\overline{\tau}_l, 1-\underline{\tau}_l]$

where $T_t = \tau_{kt} F_{kt} K_t + \tau_{lt} F_{lt} L_t$

Proof: Suppose (1)-(4) characterize a CE. Then set $\varphi^i = [\lambda_t^i]^{-1}[\beta^t \frac{\alpha}{C_t}]$ and $[c_t^i, \theta^i - l_t^i] = \varphi^i [C_t, 1 - L_t]$. Then using these expressions, we get (2). Moreover, using these expressions in (3) we get the expression φ^i in (3'). It remains to show that $\sum_i \mu_i \varphi^i = 1$. The construction of prices and taxes from the foc's(which are satisfied using φ^i) gives:

$$(1 - L_t)(1 - \tau_{lt}) F_{lt} = \frac{1 - \alpha}{\alpha} C_t$$

Then:

$$\begin{aligned}
C_t &= \alpha[(C_t - F_{lt} L_t) + \tau_{lt} F_{lt} L_t + (1 - \tau_{lt}) F_{lt}] \\
C_t &= \alpha[(-K_{t+1} + F_{kt} K_t) \pm \tau_{kt} F_{kt} K_t + \tau_{lt} F_{lt} L_t + (1 - \tau_{lt}) F_{lt}] \\
\beta^t &= \beta^t \frac{\alpha}{C_t}[-K_{t+1} + (1 - \tau_{kt}) F_{kt} K_t + T_t + (1 - \tau_{lt}) F_{lt}]
\end{aligned}$$

Notice that we have used the fact that $T_t = \tau_{kt} F_{kt} K_t + \tau_{lt} F_{lt} L_t$. Since the final expression above is true in each period, we have: [16]

$$\frac{1 - \beta^{T+1}}{1 - \beta} = \frac{\alpha}{C_0}(1 - \tau_{k0}) F_{k0} K_0 + \sum_{t \geq 0}^{T} \beta^t \frac{\alpha}{C_t}[(1 - \tau_{lt}) F_{lt} + T_t]$$

or

$$1 = \frac{(1 - \beta)}{1 - \beta^{T+1}} \left[\frac{\alpha}{C_0}(1 - \tau_{k0}) F_{k0} K_0 + \sum_{t \geq 0}^{T} \beta^t \frac{\alpha}{C_t}[(1 - \tau_{lt}) F_{lt} + T_t] \right]$$

But this is equivalent to $\sum_i \mu_i \varphi^i = 1$.

[16] We also use both $\left[\frac{\alpha \beta^t}{C_t} - \frac{\alpha \beta^{t+1}}{C_{t+1}}(1 - \tau_{k_{t+1}}) F_{k_{t+1}}\right] K_{t+1} = 0$ and $K_{T+1} = 0$

The converse is clearly true by setting $\lambda_t^i = \frac{\alpha\beta^t}{\varphi^i C_t}$ ∎

The next Lemma shows that in any point in time the wealth of individual i is an affine function of his initial share θ^i.

Lemma 8. *If $k_0^i = K_0 + (\mathcal{S}_0 - K_0)\theta^i$, in any continuation of a competitive equilibrium, $a_t^i = K_t - \mathcal{S}_t + \mathcal{S}_t\theta^i$ for some $\mathcal{S}_t \in \mathbb{R}$.*

Proof: The necessary and sufficient conditions in lemma 7 generate the following wealth levels:

$$\frac{\alpha\beta^t}{C_t}(1-\tau_{kt})r_t a_t^i = \frac{\beta^t}{1-\beta}\varphi^i - \sum_{s=t}^{T}\beta^s\frac{\alpha}{C_s}[(1-\tau_{ls})F_{ls}\theta^i + T_s]$$

Since $k_0^i = K_0 + (\mathcal{S}_0 - K_0)\theta^i$, statement (3') in lemma (7) implies that φ^i is affine in θ^i. Therefore, from the expression above, a_t^i is affine as well . Because $\int_\Theta a_t^i f(\theta^i)d\theta^i = K_t$, we have that $a_t^i = K_t - \mathcal{S}_t + \mathcal{S}_t\theta^i$ for some $\mathcal{S}_t$ ∎

For the next lemma, it is worth to explicitly express φ^i as a function of the aggregates and the type itself. Given a interior CE, let $Z_t \equiv \{C_t, K_t, L_t\}$. Then we can express the individual share for type i as $\varphi(\{Z_t, \tau_t\}_{t\geq 0}, k_0^i)$.

Define:

$$\varphi(\{Z_t,\tau_t\}_{t\geq n}, a_n^i) \equiv \frac{(1-\beta)}{1-\beta^{T+1}}\left[\frac{\alpha}{C_n}(1-\tau_{kn})F_{kn}a_n^i + \sum_{t\geq n}^{T}\frac{\alpha\beta^{t-n}}{C_n}((1-\tau_{ln})F_{ln}\theta^i + T_n)\right]$$

12 Consensus Result and DCW Equilibrium

Using proposition 1 and lemma 8, we have the following characterization of dynamic

Condorcet winner, which extends the consensus result when society does not have a commitment device.

Proposition 4. *(**Consensus Result with Non-Commitment**) Assume that $k_0^i = K_0 + (\mathcal{S}_0 - K_0)\theta^i$ and that preferences are given by (2). Take a pair $(\pi, \mathcal{A})$ consisting of a allocation mapping and a FP program. Then $(\pi, \mathcal{A})$ is a DCW if and only if:*

1. *$\forall\ t, h^{t-1} \in H_{t-1}$ with respective $\Delta\mathbf{a}_t$, and for all $\hat{\tau}_0 \in [-\underline{\tau}_l, \overline{\tau}_l] \times [0, \overline{\tau}_k]$, let $(\hat{\tau}, \hat{q})$ be generated by $\sigma|_{(h^{t-1}, \hat{\tau}_0)}$. Then $(\hat{\tau}, \hat{q})$ is a competitive equilibrium.*

2. *$\forall\, t, h^{t-1} \in H_{t-1}$, and $\forall\ \widehat{\pi}|_{h^{t-1}} \in \Pi$:*

$$U^m(\Delta\mathbf{a}_t, \sigma|_{h^{t-1}}) \geq U^m(\Delta\mathbf{a}_t, \widehat{\pi}|_{h^{t-1}}, \mathcal{A}|_{h^{t-1}})$$

where $U^m(\Delta\mathbf{a}_t, \sigma|_{h^{t-1}})$ is the continuation utility attained by the median household (θ^m) under the allocation generated by $\sigma|_{h^{t-1}}$.

Proof: (If) Consider some arbitrary history h^{t-1}. The outcome path of $\sigma|_{h^{t-1}}$ is a competitive equlibirum and, by (8), it yields a distribution of wealth $\Delta\mathbf{a}_t$ which is affine in θ_0^i. Given $\Delta\mathbf{a}_t$,then by the single crossing property of proposition 1 , if $U^m(\Delta\mathbf{a}_t, \sigma|_{h^{t-1}}) \geq U^m(\Delta\mathbf{a}_t, \widehat{\pi}|_{h^{t-1}}, \mathcal{A}|_{h^{t-1}})$ we have that $U^i(\Delta\mathbf{a}_t, \sigma|_{h^{t-1}}) \geq U^i(\Delta\mathbf{a}_t, \widehat{\pi}|_{h^{t-1}}, \mathcal{A}|_{h^{t-1}})$ either by all $\theta^i \leq \theta^m$ or all $\theta^i \geq \theta^m$.
This implies $(\pi|_{h^{t-1}}, \mathcal{A}|_{h^{t-1}}) \succeq_M (\widehat{\pi}|_{h^{t-1}}, \mathcal{A}|_{h^{t-1}})$.

Since the history h^{t-1} was arbitrary, conditions 1 and 2 in the statement imply that $(\pi, \mathcal{A})$ is a DCW.

(Only If) Let $\{Z_s, \tau_s\}_{s\geq t}$ be the outcome path of taxes and aggregates generated by $\sigma|_{h^{t-1}}$ and $\{\hat{Z}_s, \hat{\tau}_s\}_{s\geq t}$ the respective outcome generated by $\widehat{\pi}|_{h^{t-1}}, \mathcal{A}|_{h^{t-1}}$.Suppose that $(\pi, \mathcal{A})$ is a DCW and that condition 2 does not hold at some history h^t:

$$U^m(\Delta \mathbf{a}_t, \sigma|_{h^{t-1}}) < U^m(\Delta \mathbf{a}_t, \widehat{\pi}|_{h^{t-1}}, \mathcal{A}|_{h^{t-1}})$$

This can be rewritten as:

$$\log(\varphi(\{Z_s, \tau_s\}_{s\geq t}, a_t^m)) - \log(\varphi(\{\hat{Z}_s, \hat{\tau}_s\}_{s\geq t}, a_t^m)) > V(\{\hat{Z}_s, \hat{\tau}_s\}_{s\geq t}) - V(\{Z_s, \tau_s\}_{s\geq t})$$

By the single-crossing property, since the median type prefers one to the other, more than the majority supports the alternative path, which contradicts

$(\pi|_{h^{t-1}}, \mathcal{A}|_{h^{t-1}}) \succeq_M (\widehat{\pi}|_{h^{t-1}}, \mathcal{A}|_{h^{t-1}})$ ∎

Proposition 4 simplifies to a great extent the task of checking the requirements for a DCW. In particular, condition 2 in the definition of a DCW can be replaced without loss of generality in this economy by the second condition in Proposition 4. Therefore the condition that the continuation $\pi|_{h^t}$ defeats by majority any other plan, given the individual decisions of agents, is replaced by a time-consistent requirement. Moreover, time-consistency is evaluated at the median's allocation on the outcome path given the distribution of wealth after any history of taxes.

As a byproduct of proposition 4, DCW equilibrium in the economy considered here is equivalent to sustainable equilibrium (as in Chari and Kehoe (1990) and Phelan and Stacchetti (2001)) when the government sets redistributive fiscal policy and assigns welfare weight equal to one to the median type.

Part IV

Heterogeneous Beliefs and Optimal Taxation[17]

13 Introduction

Since the seminal work of Kydland and Prescott (1977), a great deal of research has been devoted to the study of time inconsistency problems. In short, an optimal policy under commitment is time inconsistent if its continuation plan is not optimal. In another words, governments or policy makers have the incentive to break their promises when re-optimizing ex post, giving rise to inefficient outcomes. These types of problems have been studied in a wide variety of policy settings but the most common ones are related to capital taxation, optimal monetary policy and default decisions. In their original work, Kydland and Prescott (1977) propose that the way to avoid these problems is to tie the hands of policy makers after the fact by forcing them to use rules (set ex ante with commitment) rather than allowing them discretion (set ex post, without commitment). The difficulty with this approach is that it undermines the capability of the government to react optimally when there are changes in the fundamentals.

Following up on this idea, Chari and Kehoe (1990) and Chari and Kehoe (1993)

[17] This chapter is coauthored with **Facundo Piguillem.**

showed that if the policy maker is patient enough, the optimal policy under commitment is sustainable even when the government is endowed with full discretion. Their argument relies on reputational considerations borrowed from game theory. A key factor in the result is the assumption that information is complete and perfect. As a consequence of this assumption, if after some history the government is found to have deviated from the optimal policy, every agent would observe it and all of them would, in the best equilibrium, coordinate to punish the government. Notice that to achieve this kind of equilibrium it is important not only that all agents know that the government has deviated but also that all agents know that everyone knows, that everyone knows and so on–the usual recursion due to common knowledge. Therefore, coordination is not only possible but also perfect. Because of their simplicity and tractability these kinds of models have become the dominant tool to analyze environments without commitment in macroeconomics. Thus, although time inconsistency problems play an important role in public policy, reputational considerations can be used as a way of solving the problem without the need of institutional reforms designed to mimic commitment.

This paper analyzes time inconsistency problems in economies with incomplete information. Specifically, we study an environment similar to Chari and Kehoe (1990) in which an investment decision must be made by private agents before the government sets a capital income tax. In that setting, if there is no commitment and a finite horizon, governments (even benevolent ones) will set capital tax rates too high. They show however, that in an infinitely repeated setting, trigger strategies can enforce the commitment outcome when agents are sufficiently patient. We deviate from the Chari and Kehoe example in two important ways. First, we assume that aggregate productivity in the economy (the fundamental) is stochastic. This by itself does not change the Chari and Kehoe result - if agents are patient enough, the full commitment outcome can be supported. Second, we assume that agents do not share the same information. Specifically, we assume that the government sees the true aggregate state, but that this is not observed by private agents. Rather, each agent privately receives a different signal (payoff relevant) about the aggregate state of the economy. The signals can be made arbitrarily close to the aggregate state.

Here, agents have different information sets, and therefore different beliefs, about the fundamentals. In this environment the optimal policy under commitment depends solely on the state of the fundamental. However, if the government deviates agents cannot be certain if what happened is actually a deviation or is the optimal reaction to a change in the fundamentals. In addition, agents do not know other agents' beliefs. As a result, independent of the accuracy of the signal, incomplete information reduces the set of equilibrium payoffs.

First, we show that strategy profiles for the government that depend solely on the fundamentals along the equilibrium path cannot be equilibrium profiles. In particular, there is no strategy profile in the repeated game that delivers the best allocation under commitment, regardless the punishment prescription off the equilibrium path. Second, we show that when government's private shock takes on two values and agents are patient enough the best equilibrium can be achieved with strategy profiles that depend only on public histories. In another words, the best equilibrium is a policy independent of the fundamentals. Finally, for any discount factor strictly smaller than one the best equilibrium is inefficient.

The first result is in line with the literature about repeated games with private imperfect monitoring and a finite number of players (e.g. Mailath and Samuelson (2006)). However, the reasoning behind the argument is slightly different. In games with private imperfect monitoring problems arise when players do not know exactly in which state they are (on or off the equilibrium path) and therefore they are not able to coordinate to punish each other off the equilibrium path. In our environment, the government always knows with certainty everything that has happened while the agents have different beliefs about past histories of the fundamental. If the agents trusted the government when using a strategy that depends only on its private information the government would defect and no agent would be able detect the deviation. Thus, any equilibrium strategy has to depend on some object that is fully observed by the agents. Since in our environment the only variable that is perfectly observed by every player is the tax on investment (the action space for the government) any equilibrium strategy has to depend on past taxes. Moreover, if the strategy for the government depends only on the history of taxes then the results about environments with perfect and complete information carry over entirely. The fact that the government's strategy depends on a history that is perfectly observed by every player allows the full coordination of the agents to punish the government if it deviates. Although the size of the set of public equilibria depends on the size of the discount factor, in the economy analyzed here this kind of strategy generates payoffs that are uniformly bounded away from the best one (the best payoff under commitment).

Would equilibrium strategies that depend on both public and private information increase the payoff? The answer is no. Given that agents have no way to foresee the fundamental, the game becomes one of repeated adverse selection. Thus, when an equilibrium strategy depends on private information punishments happen with positive probability on the equilibrium path. The punishment takes the form of a smaller continuation payoff after those actions that are especially tempting. Nonetheless, the "punishment cost" could be compensated with a larger present payoff fitting the present action to the realization of the fundamental. Unlike the usual environments in game theory, here the agent and the principal have the same payoff functions. Therefore, optimal punishments along the Pareto frontier arbitrarily close to the optimal average welfare are not available. In the language of contract theory, every punishment for the agent (the government) would hurt the principal as well (the agents in this economy). Since the punishment cost is always at least as large as the gain from discretion, the best equilibrium implies a policy that is independent of the fundamental.

Regarding the literature about this topic, to the best of our knowledge there are three closely related papers: Sleet (2001), Athey, Atkeson, and Kehoe (2005) and Sleet and Yeltekin (2006). The first paper considers the problem of a monetary authority that receives a private signal about the true state of the economy and both households and firms have the same information set. They show that under some conditions the optimal policy with commitment is an equilibrium, while in other cases the monetary authority chooses not to use the private signal. The second paper, again in an optimal monetary policy context, considers an environment where agents have the same information sets and only the policy maker observes the (random) true state of the economy. They find that if the time inconsistency problem is "severe" the optimal policy is independent of the true state of the economy; otherwise some dependency is allowed. Sleet and Yeltekin (2006) also analyze an economy with government debt in which private agents have the same information sets, but the government privately observes a taste shock related to the public good consumption. They find that the interaction between informational frictions and the possibility of debt repudiation yields more persistence in both taxes and debt (in the best sustainable allocation) when compared to the benchmark economy (with full commitment and complete information).

This paper differs from the above in the following ways. First, agents have different information sets. Second, in the above papers the information privately known by the government does not directly affect either the payoff or the feasibility set of the agents. Consequently, agents cannot extract from their information sets any useful information about the signals received by the government. In this paper, agents can foresee in an arbitrarily precise way the signal received by the government. Therefore, their results can be viewed as the limit case of the economy studied here. Finally, all the papers mentioned before analyze Public Perfect Equilibria. That is, the strategy space of the government is constrained to include only the last realization of its private information. This paper extends the result to the unconstrained strategy space.

The paper proceeds as follows: Section 16 describes the environment. Section 17 defines and characterizes the equilibria. In Sections 18 and 19 we characterize the best equilibrium under commitment. Section 20 shows the inefficiency result. Section 21 describes an alternative environment for which the results go through. The last section concludes.

14 The Economy

14.1 Uncertainty

We consider a repeated game with a benevolent government and a continuum of households indexed by $i \in I = [0, 1]$. Time is discrete and indexed by $t = 0, 1, 2, ...$ At the the beginning of every period t, the outcome $R_t \in \Upsilon$ of a random variable

$\hat{R}_t$ is realized. The set Υ has cardinality equal to N. Let $\bar{R} \equiv max_{\hat{R}\in\Upsilon}\{R\}$. The outcome R_t is observed only by the government. Each $\hat{R}_t$ is distributed i.i.d. over time with probability distribution P. The process $\{\hat{R}_t\}_{t=0}^{\infty}$ is independent of any choices made by the government or households.

In each period, conditional on the realization of $\hat{R}_t$, each individual privately observes a draw $y_t^i \in [1, \overline{y}]$ of a random variable Y from the probability density function $f(y|R_t)$ and distribution function $F(y|R_t)$. The value $\overline{y}$ may be infinite. Conditional on the realization of the aggregate state, the individual shocks are i.i.d. across agents. We assume a version of the Law of Large Numbers with a continuum of random variables relying on the construction of Sun (2006). We also shall impose the following.

Assumption 1. *Stochastic processes.*

1. *For all* $R, R' \in \Upsilon$, $F(y|R) > F(y|R')$ *if* $R' > R$. *(First order stochastically dominance)*

2. $f(y|R) > 0$ *for all* $R \in \Upsilon$ *and almost all* $y \in [1, \overline{y}]$. *(Full support)*

3. *For all* $R \in \Upsilon$, $\int_1^{\overline{y}} y f(y|R) dy = R$

The first condition assures that given individual actions, aggregate output and aggregate investment are strictly increasing in the realizations of the aggregate state. The second condition is a technical assumption, standard in the literature, that prevents dealing with zero probability events. The full support assumption makes sure that everything can happen in every period, independently of the realization of $\hat{R}$. Finally, the third assumption is just for simplicity and to save on notation.

14.2 Stage Game: Actions and Payoffs

There are three goods in the economy in each point in time: two private goods, consumption and investment, and a public good. In each period every agent receives a physical endowment of $\omega > 0$ units of the private good. After observing the individual shock y^i, each household chooses investment $x_t^i(y^i) \in [0, \omega]$. Returns on investment in period t are agent-specific, and given by the draw y_t^i.

Given individual investment decisions, the realized aggregate output, $Y(R_t) = \int_0^1 \int_1^{\overline{y}} y_t^i x_t^i(y_t^i) f(y_t^i|R_t) dy_t^i di$ and aggregate investment $X(R_t) = \int_0^1 \int_1^{\overline{y}} x_t^i(y_t^i) f(y_t^i|R_t) dy_t^i di$ are only observed by the government. Next, the government chooses a tax on investment $\tau_t \in [0, 1]$. Agents are risk neutral in the private good. If household i invests x_t^i and the government sets taxes equal to τ, then its individual consumption is given by $c^i = (1 - \tau) y^i x^i + (\omega - x^i)$.

There is a technology that automatically transforms aggregate output into a public good g. If the aggregate state is R_t and the government sets tax on investment τ_t then the amount of public good provided, as a function of the aggregate state, is given by:

$$g_t(R_t) = \tau_t Y_t(R_t) \tag{21}$$

We assume that individual agents do not observe the provided amount of the public good. This is a strong assumption that greatly simplifies the analysis and the exposition of the main results. As we explain in Section 19 there is an alternative environment where the government does not observe R but a signal related to it, and both the agents and the government observe the all aggregates in the economy at the end of each period. As long as the signal is not perfectly informative the main results of this paper are still true.

Preferences are separable between the private and the public good. If household i gets a draw y^i, invest x^i, the tax on investment is τ and aggregate output is Y, its payoff is given by:

$$u(y^i, \tau, Y, x^i) = [(1-\tau)y^i x^i + (\omega - x^i) + v(\tau Y)] \tag{22}$$

where $v : \mathbb{R}_+ \mapsto \mathbb{R}$ is twice continuously differentiable and strictly concave function.

Given the realization of R_t, the profile of individual investment functions and tax τ_t, government's payoff is given by:

$$W(R_t, \tau_t; x^i_{i \in I}) = \int_0^1 \int_1^{\overline{y}} u(y^i_t, \tau_t, Y_t, x^i_t) f(y^i_t | R_t) dy^i_t di$$

We abuse notation writing the above function as in (23) below. We can do this because given any profile of individual investment functions the payoff for the government depends only on the aggregate values for investment and output. In addition, as we show later since agents are ex-ante identical, the decision function for all agents are equal.

$$W(R_t, \tau_t, X_t, Y_t) = \int_1^{\overline{y}} u(y_t^i, \tau_t, Y_t, x_t^i) f(y_t^i | R_t) dy_t^i \tag{23}$$

The following sequence of events summarizes the information structure of the stage game:

1. R is realized;
2. y^i is drawn for each i;
3. Each individual chooses x^i;
4. Government privately observes R_t and sets τ;
5. The public good is produced according to (21);
6. Consumption is realized.

In order for time inconsistency to play a role, we shall assume that the marginal value of provision of the public good is higher than the marginal value of individual consumption in any possible state.

Assumption 2 (**Time Inconsistency**). *$v'(\bar{R}) > 1$.*

To understand this condition, take any period t after which all the agents have chosen x_t^i, and therefore both R_t and y_t^i have been realized. Let X_t be the aggregate investment at time t. Then, using the government budget constraint, the one period utility for the government is given by:

$$W(R_t, \tau_t, X_t, Y_t) = (1-\beta)[(1-\tau_t)Y_t + (\omega - X_t) + v(\tau_t Y_t)]$$

If the government increases taxes slightly, say by $d\tau_t$, the benefit of increasing the public good is given by $Y_t(R_t)v'(\tau_t Y_t(R_t))d\tau_t$, while the loss in private consumption is given by $Y_t(R_t)d\tau_t$. Combining these two effects, the government has incentives to increase the current tax as long as $Y_t(R_t)[v'(\tau_t Y_t(R_t)) - 1] > 0$, which is guaranteed by Assumption 2.

15 Perfect Bayesian Equilibrium

In the repeated game, a public history is a collection of variables that have been observed by all the players. At the beginning of period t, a public history is $h^{P,t} \equiv \{\tau_0, ..., \tau_{t-1}\}$. Let the set of all public histories $h^{P,t}$ at time t be given by $H^{P,t}$. In contrast, a history for the government at time t consists only of the observed outcomes by the government. In terms of outcome paths, because households are competitive there is no loss of generality if we define private histories for the government which do not include the aggregates[18]. In this way, let $h^{g,t} \equiv \{h_0^g, ..., h_t^g\}$ $\in H^{g,t}$ with $h_s^g = \{\tau_{s-1}, R_{s-1}\}$ if $t > 1$ and $H^{g,0} = \emptyset$. A history for individual i is given by $h^{i,t} \equiv \{h_0^i, ..., h_t^i\} \in H^{i,t}$ with $h_s^i = \{\tau_{s-1}, y_{s-1}^i\}$ if $s > 1$ and $H^{i,0} = \emptyset$.

[18]For a detailed explanation of this reasoning, see Chari and Kehoe (1990)

The information sets for an individual player i at t correspond to all histories of the game $h^t \equiv \{h_0, ..., h_t\} \in H^t$ with $h_s = \{\tau_{s-1}, R_{s-1}, \{y^i_{s-1}\}_{i\in I}\}$ that are consistent with her own history at time t.

We restrict the analysis to pure strategies for the government. A pure strategy for the government is a sequence $\{\sigma_{G,t}\}_{t=0}^{\infty}$ with $\sigma_{G,t} = H^{g,t} \times \Upsilon \to [0,1]$. A strategy for an agent $i \in I$ is given by $\{\sigma_{i,t}\}_{t=0}^{\infty}$ with $\sigma_{i,t} : H^{i,t} \times [1, \bar{y}] \to [0, \omega]$. Both $\sigma_{G,t}$ and $\sigma_{i,t}$ are assumed to be measurable functions.

In order to consider any kind of perfection, given the informational restrictions, individual agents have to form beliefs over their information sets. We have opted to analyze Perfect Bayesian equilibria. Let $\mu(\cdot|\tilde{h}^{i,t}, y^i)$ be the probability distribution over histories $\hat{h}^{g,t} \in H^{g,t}$ consistent with individual history $\tilde{h}^{i,t}$. Let Σ_G be the set of possible strategy profiles for the government and Σ be the set of possible strategy profiles $\sigma = (\sigma_G, \{\sigma_i\}_{i\in I})$. A strategy profile $\sigma \in \Sigma$ induces, after any history $h^t \in H^t$, a continuation profile $(\sigma_G|_{h^{g,t}}, \{\sigma_i|_{h^{i,t}}\}_{i\in I}) \in \Sigma$.

Given the risk neutrality assumption and the fact that agents are both anonymous and atomistic, optimality for the individuals can be reduced to a simple rule. Given $\sigma_G \in \Sigma_G$, let $E_{\sigma_G}(\tau|h^{i,t}, y^i)$ be the conditional expectation that a household with history $h^{i,t}$ and idiosyncratic return y^i has about the random variable $\sigma_{G,t}$. This function, for each individual i, is measurable with respect to the sigma-algebra generated by his individual histories. A household with history $h^{i,t}$ and idiosyncratic return y^i will invest a positive amount only if the expected marginal return on investment is positive:

$$x^*(h^{i,t}, y^i) = \begin{cases} \omega & \text{if } y^i E_{\sigma_g}(1-\tau|h^{i,t}, y^i) - 1 > 0 \\ [0, \omega] & \text{if } y^i E_{\sigma_g}(1-\tau|h^{i,t}, y^i) - 1 = 0 \\ 0 & \text{otherwise} \end{cases} \tag{24}$$

Given a profile $\sigma \in \Sigma$ and a sequence of belief profiles $\mu \equiv \left\{(\mu(\cdot|h^{i,t}, y^i))_{i \in I}\right\}_{t=0}^{\infty}$, expected payoffs for the players are naturally defined from the stochastic outcomes that the strategies induce. The payoff for the government at time zero in the repeated game is given by:

$$V(\sigma) = (1-\beta)E\left[\sum_{t=0}^{\infty} \beta^t W(R_t, X_t, \tau_t, Y_t)\right]$$

Notice that, since the government observes all the aggregates in the game, it does not need to form beliefs about individuals' actions. It only needs to take into consideration that its own actions affect individuals' beliefs.

Definition 6. *A pair (σ, μ) consisting of strategy profiles and belief profiles is a Perfect Bayesian Equilibrium if:*

(i) Given $\{\mu^i\}_{i \in I}$, $\sigma_{i,t}(h^{i,t}, y^i) = x^*(h^{i,t}, y)$ $\forall i \in I, h^{i,t} \in H^{i,t}, \forall y^i \in [1, \bar{y}]$;

(ii) $V(\sigma_G|_{h^{g,t}}, \{\sigma_i\}_{i \in I}) \geq V(\widetilde{\sigma}, \{\sigma_i\}_i)$ $\forall \widetilde{\sigma} \in \Sigma_G, h^{g,t} \in H^{g,t}, \forall R$;

(iii) Beliefs are given by Bayes' rule whenever possible.

Conditions (i) and (ii), respectively, require that, given beliefs, the government's and the individual's continuation strategies be best responses to each other after any history. Households' deviations cannot be detected and therefore at each period they maximize, given their beliefs, utility from private consumption. Regarding individuals' decisions about investment, we shall assume the following:

Assumption 3 (**Monotone Likelihood Ratio**). *For all* $R_H, R_L \in \Upsilon$ *and for all* $\hat{y}, y \in [1, \bar{y}]$ *we have that* $\frac{f(\hat{y}|R_L)}{f(y|R_L)} \leq \frac{f(\hat{y}|R_H)}{f(y|R_H)}$ *if* $\hat{y} \geq y$ *and* $R_H > R_L$.

Assumption 4 (**Analytical pdf**). *For each* $R \in \Upsilon$, $f(.|R) : [1, \bar{y}] \to \mathbb{R}$ *is analytic.*

A function defined on the real line is analytic if it is equal to its Taylor expansion. The role of Assumption 4 is to guarantee that for any $\sigma \in \Sigma$, the set of individuals indifferent between investing or not has Lesbegue measure zero. This assumption is not crucial for the main result of this paper but simplifies the characterization of the individual decisions. Without it, is not even clear that aggregate investment, and therefore government's revenues, are decreasing in individual taxes. Analytical functions are fairly common among continuous and differentiable distributions.[19]

Lemma 9. *Under Assumption 4, for any* $\sigma \in \Sigma$ *and any* $h^t \in H^t$, *the set of individuals for which* $y^i E_{\sigma_g}(1 - \tau | h^{i,t}, y^i) - 1 = 0$ *in (24) has Lesbegue measure zero.*

[19] Fox example, Normal, Uniform, Exponential, Gamma, Beta and Pareto are analytic. Among the distributions satisfying the monotone likelihood ratio are the Normal, Exponential, Uniform and Beta.

Proof: In the appendix.

Assumption 3 is a technical condition that we use to deal with the case in which there are multiple agents indifferent between investing or not.

16 Ramsey Equilibrium

Before proceeding with the characterization of Perfect Bayesian equilibria, we first consider the benchmark case in which the government has a commitment technology that it is used to bind itself to a tax policy $\sigma_G : \Upsilon \to [0, 1]$ in each period. When such technology is available, the static nature of the government's problem allows us to restrict the analysis to an one-period game. Following the literature, we call it the Ramsey game. The introduction of a commitment technology can be formalized by changing the timing of the one shot Bayesian game. The Ramsey game that we analyze evolves as follows:

0. The government sets a tax policy $\sigma_G : \Upsilon_R \to [0, 1]$;
1. R is realized;
2. y^i is drawn for each i;
3. Each individual chooses x^i;
4. Government observes both R and aggregate output Y and it sets τ according to $\sigma_G(\cdot)$;

5. Public good is produced according to equation (21);

6. Consumption of both goods is realized.

There are two differences with the stage game in Section 2. First, the government sets a tax policy $\sigma_G(\cdot)$ before observing R. Second, only after aggregate output is realized government learns about the realization of R and sets the investment tax according to $\sigma_G(\cdot)$. This specific choice about the sequence of events implies the existence of a strategy profile in the Bayesian game that attains, despite incentive questions, the outcome in the Ramsey game. At the same time it prevents any discussion about communication issues.

Given a a tax policy σ_G, let $E_{\sigma_G}(\tau|y^i)$ be the conditional expectation that a household with draw y^i has about the random variable σ_G .

Definition 7. *The Ramsey equilibrium is a function* $\sigma_G : \Upsilon_R \to [0,1]$ *and, for each* $i \in [0,1]$, *a function* $\sigma_i : [1, \overline{y}] \to [0, \omega]$ *such that:*

a) σ_G *maximizes* $\int_\Upsilon W(R, \tau, Y) dP(R)$ *given* $\sigma_i|_{i \in I}$.

b) $$\sigma_i^*(y) = \begin{cases} \omega & \text{if } y \cdot E_{\sigma_G}(1-\tau|y) - 1 > 0 \\ [0,\omega] & \text{if } y \cdot E_{\sigma_G}(1-\tau|y) - 1 = 0 \\ 0 & \text{otherwise} \end{cases}$$

It turns out to be difficult to characterize precisely the Ramsey policy with a general function $v(.)$. All what we can say is that $g(R) < g(R')$ if $R < R'$, that is, the government spending is higher in more productive states.[20] Since aggregate output, $Y(R)$, is increasing in R as well, the magnitude of the taxes is unclear. The properties and relative size of taxes would be very different depending on the shape of $v(.)$, even when it is assumed that this function is strictly concave and twice differentiable. For that reason in the next Lemma we assume that $v(.)$ is linear. That is, $v(x) \equiv b \cdot x$. Of course, because of Assumption 2, $b > 1$. When that is the case the characterization is intuitive and straightforward. Moreover, it highlights the main complications related to the individual investment decisions. The next lemma characterizes the equilibrium of the Ramsey game.

Lemma 10. *Under Assumptions 2-4 if $v(.)$ linear, then the Ramsey equilibrium is given by:*

1. $(1 - \sigma_G^*(\bar{R}))\sigma_G^*(R) = 0$ for all $R \in \Upsilon$, $R \neq \bar{R}$
2. $\sigma_i^*(y^i)$ is given by item b) in the definition of the Ramsey equilibrium.

Proof: In the appendix.

Lemma 10 states that the solution to the Ramsey game with linear utility is in a corner. By Assumption 2, taxes being either zero or one in all states cannot be a solution. Moreover, the government taxes a positive amount in the highest aggregate state. This happens because taxing in the highest state is always less

[20] Here the difficulty is similar as in the Mirrleesian literature, where is not known in general whether or not workers with higher productivity work more than low productivity workers.

costly than taxing in a lower state. For any given average tax, it is always possible to increase the payoff by increasing the tax in the highest state and reducing the tax in a lower state in such a way that the average tax remains the same. The proof of the lemma exploits this idea. It is worth to comment the role of Assumptions 3-4 in the lemma. If the function $H(y;\tau) = y \cdot E_{\sigma_G}(1-\tau|y) - 1$ had the single crossing property[21], then the characterization of the individual decisions would be very simple. This is not always the case for any possible tax function $\tau : \Upsilon \to [0,1]$. In the proof of Lemma 10 we exploit those assumptions in order to handle the more general case in which there are multiple agents indifferent between investing or not.

For future reference, we define the Ramsey strategy profile $(\sigma_G^*, \sigma_i^*) \in \Sigma$ as the repetition of the Ramsey equilibrium. Formally, for all $h^{g,t} \in H^{g,t}$ and $\hat{R} \in \Upsilon$, let $\sigma_{G,t}^*(h^{g,t}, \hat{R}) = \sigma_G^*(\hat{R})$. Regarding household's strategies, for all $h^{i,t} \in H^{i,t}$ and $y^i \in [1, \overline{y}]$, let $\sigma_{i,t}(h^{i,t}, y^i) = \sigma_i^*(y^i)$ in Lemma 10. Beliefs are given by $\mu(\hat{R}|\hat{y}^i) = \frac{f(\hat{y}^i|\hat{R})P(\hat{R})}{\sum_{R\in\Upsilon} f(\hat{y}^i|R)P(R)}$.

17 Unattainability of the Ramsey Outcome

In this section we show that there is no strategy profile $(\sigma_G, \sigma_i) \in \Sigma$, together with some belief system, that yields the outcome path of the Ramsey equilibrium. If the Ramsey outcome were an equilibrium, it should be the case that **on the equilib-**

[21]By single crossing we mean the following: there exists $y^* \geq 1$ such that $\sigma_i(y) = \omega$ if $y \geq y^*$ and $\sigma_i(y) = 0$ otherwise.

rium path the government was playing actions that depend only on the current shock R_t. But then, the government would have profitable deviations. For instance, every time that the government is supposed to play the lowest tax prescribed by the equilibrium strategy it could choose the highest tax consistent with equilibrium behavior. Because agents cannot be certain about the real value of R this deviation would be undetectable. This result remains true regardless the punishment prescription off the equilibrium path. In order to state proposition 5, given a pair of strategies $\sigma = (\sigma_g, (\sigma_i)_{i \in I}) \in \Sigma$ and a belief profile μ, the outcome path is denoted by a sequence of stochastic functions $Z = \{(x_t^i(\sigma), c_t^i(\sigma))_{i \in I}, g_t(\sigma), \tau_t(\sigma)\}_{t=0}^{\infty}$. As usual, the outcome path is defined as the induced outcome starting from the initial history $H^{g,0}$.

In what follows, let $Z^* = \{\{x_t^i(\sigma^*), c_t^i(\sigma^*)\}_{i \in I}, g_t(\sigma^*), \tau_t(\sigma^*)\}$ denote the outcome path of the Ramsey allocation, where σ^* was defined at the end of the last section.

Proposition 5. *Under assumptions 2 and 4 there is no belief system μ and $\sigma \in \Sigma$ that generates Z^* on the equilibrium path.*

Proof: Suppose not, that is, there is a pair of strategy profiles $\hat{\sigma} \in \Sigma$ and a belief system μ such that Z^* is an equilibrium outcome.

Let $S = \left\{ \tau \in [0,1] : \tau = \sigma^*(R_t), R_t \in \Upsilon \right\}$ and define:

$$H^{*,i} = \{h^{i,t} \in H^i : \tau_s \in S, \forall s \leq t-1\}$$

Notice that $H^{*,i}$ is the set of possible histories on the equilibrium path for household i. In addition, because of the full support assumption, all individual histories in $H^{*,i}$ have non-zero measure.

Given $\hat{\sigma} \in \Sigma$, let $\mu(\cdot|h^{i,t}, y_t^i)$ be the induced probability distribution over $H^{g,t} \times \Upsilon$ given the history $(h^{i,t}, y_t^i)$. In the appendix we show that, for all $h^{i,t} \in H^{*,i}$, the belief system should follow the following updating:

$$\mu(h^{g,t}, R_t|\tilde{h}^{i,t}, y_t^i) = P(R_t|y_t^i)\frac{\mu(h^{g,t-1}|\tilde{h}^{i,t-1})}{\sum_{\hat{h}^{g,t-1}}\mu(\hat{h}^{g,t-1}|\tilde{h}^{i,t-1})}$$

By recursive calculations the above expression can be reduced to

$$\mu(h^{g,t}, R_t|\tilde{h}^{i,t}, y_t^i) = f(y_t^i|R_t)\prod_{s=0}^{t-1}\frac{f(y_s^i|R_s)}{\sum_{\hat{R}_s} f(y_s^i|\hat{R}_s)} \tag{25}$$

That is, belief are unaffected by government actions.

In addition, $\mu(\hat{\sigma}(h^{g,t}, R_t)|h^{i,t}, y_t^i) = \mu(\hat{\sigma}(h^{g,t}, R_t)|\tilde{h}^{i,t}, y_t^i)$ for all $h^{i,t}, \tilde{h}^{i,t} \in H^{*,i}$ and all $y_t^i \in [1, \overline{y}]$. Therefore, equation (24) implies that $\hat{x}^i(h^{i,t}, y_t^i) = x^i(\tilde{h}^{i,t}, y_t^i)$ for all $h^{i,t}, \tilde{h}^{i,t} \in H^{*,i}$ and all $y_t^i \in [1, \overline{y}]$. Take some period t and $R' \in \Upsilon$ such that $\hat{\sigma}(h^{g,t}, R') < \max_{\tau \in S} \tau$. Then consider the following one shot deviation by part of the government:

$$\tilde{\sigma}_G(h^{g,s}, R) = \begin{cases} \tau^D \equiv \max_{\tau \in S} \tau & \text{if } s = t \text{ and R=R'} \\ \hat{\sigma}(h^{g,s}, R) & \text{otherwise} \end{cases}$$

Following history $(h^{g,t}, R')$, the equilibrium strategy generates a government's payoff of:

$$(1-\beta)W(R_t, \hat{\sigma}(h^{g,t}, R'), X(h^{g,t}), Y(h^{g,t}))+$$
$$\beta V(\hat{\sigma}|_{\{h^{g,t}, R', \hat{\sigma}(h^{g,t}, R')\}}, (\hat{\sigma}_i|_{\{h^{i,t}, y_t^i, \hat{\sigma}(h^{g,t}, R')\}})_{i\in I}) \quad (26)$$

where $X(h^{g,t})$ and $Y(h^{g,t})$ are the aggregates following the outcome path after history $h^{g,t}$.[22]

The one shot deviation strategy generates a payoff of:

$$(1-\beta)W(R_t, \tau^D, X(h^{g,t}), Y(h^{g,t})) + \beta V(\hat{\sigma}|_{\{h^{g,t}, R', \tau^D\}}, (\hat{\sigma}_i|_{\{h^{i,t}, y_t^i, \tau^D\}})_{i\in I}) \quad (27)$$

Since $\tau^D \in S$ it follows that $\{h^{i,t}, \tau^D, y_t^i\} \in H^{*,i}$ and therefore $\hat{\sigma}^i|_{\{h^{i,t}, \hat{\sigma}(h^{g,t}, R'), y_t^i\}} = \hat{\sigma}^i|_{\{h^{i,t}, \tau^D, y_t^i\}}$ for all $i \in I$. In addition, $\hat{\sigma}|_{\{h^{g,t}, R', \hat{\sigma}(h^{g,t}, R')\}} = \tilde{\sigma}|_{\{h^{g,t}, R', \tau^D\}}$ by construction. Thus, the continuation payoffs are equal in both (26) and (27). By assumption 2, $\hat{\sigma}(h^{g,t}, R') < \tau^D$ implies that:

$$W(R_t, \tau^D, X(h^{g,t}), Y(h^{g,t})) > W(R_t, \hat{\sigma}(h^{g,t}, R'), X(h^{g,t}), Y(h^{g,t}))$$

. Hence the deviation is profitable, a contradiction ∎

Assumption 4 plays a role in proposition 5 only to the extent that it guarantees that a measure zero of agents is indifferent between investing or not. A crucial feature that prevents any strategy achieving the Ramsey outcome is the fact that, otherwise, the individual strategies do not depend on history on the equilibrium

[22] Since individual decisions are measurable with respect to their information sets, the aggregates cannot depend on the actual realization of R_t after history $h^{g,t}$.

path. The government then takes advantage of this situation, defecting whenever possible.

Although the Ramsey payoff cannot be attained, one may wonder if it can be approached arbitrarily close for high enough discounting. The answer to this question is negative under some circumstances, and we will elaborate it in the next section. Notice that if indeed there is a strategy profile than can approach the (repeated) payoff of the Ramsey equilibrium, such profile requires some coordination among agents. In another words, a positive measure of agents should have strategies depending on public histories. If only a measure zero of agents could coordinate any punishment that they could used would have no effect on the government's payoff.

Proposition 5 can actually be made stronger. As its proof makes clear, there is nothing special to the Ramsey outcome other than **on the equilibrium path** taxes are independent of history [23] and stochastic.

In order to formalize the above statement, consider the following class of strategies;

Definition 8. *Let $R^t = (R_{0,\ldots,}R_t) \in \Upsilon^t$ be a history of the aggregate state. A strategy profile $\sigma_G \in \Sigma_G$ is* ***purely private*** *if for all t, $h^{g,t}, \hat{h}^{g,t} \in H^{g,t}$ $\sigma_{G,t}(h^{g,t}, R_t) = \sigma_{G,t}(\hat{h}^{g,t}, \hat{R}_t)$ if $R^t = \hat{R}^t$.*

Definition 9. *A strategy profile $\sigma_G \in \Sigma_G$ is* ***non-trivially purely private*** *if*

[23] Or in another words, the outcome is just the repetition of the static Ramsey equilibrium.

1. *It is purely private;*

2. *There exist* t *and* $R^t, \hat{R}^t \in \Upsilon^t$ *such that* $R^t \neq \hat{R}^t$ *with* $\sigma_{G,t}(h^{g,t}, R_t) \neq \sigma_{G,t}(\hat{h}^{g,t}, \hat{R}_t)$.

As in Proposition 5 we can define the outcome path of a purely private strategy profile for the government in the usual way. Then, we have the following proposition;

Proposition 6. *Under Assumptions 2-4, there is no belief system* μ *such that the outcome of* ***non-trivially purely private*** *strategy profile for the government is a Perfect Bayesian Equilibrium.*

Proof: In the appendix.

The main steps of the proof are similar to those in Proposition 5. If **on the equilibrium path** the government were following a purely private strategy, there would be more than one tax consistent with equilibrium behavior in at least one period. Then, when is time to choose the low tax, the government could deviate choosing the highest tax prescribe by the strategy. By the same arguments as in Proposition 5 these kinds of deviations are not detectable. The main implication of Proposition 6 is that the problem becomes one of repeated hidden information as if the households had not information whatsoever about the true state of the economy. This fact allow us to use the usual tools for repeated agency problems where the households are the principal and the government is the agent.

18 Best Equilibrium and Inefficiency Result

Proposition 6 implies that in any equilibrium in which the government uses pure strategies, its strategy profile has to condition actions on public information somehow. If the government condition its actions only on public histories (of previous taxes), a wide range of payoffs can be sustained. In this case the private agents can fully coordinate to punish deviations made by the government, switching to the worst equilibrium. In the worst equilibrium, the government indeed uses a strategy profile that only depends on public information. Regardless the history, the government always taxes investment fully. Anticipating this behavior, agents do not save. As a consequence, government's action after any history is also a best response to individual agents' strategies.

Proposition 7 (Worst Equilibrium). *The pair of strategies* $\sigma_{G,t}^{worst}(h^{g,t}, R_t) = 1$ *for all* $h^{g,t} \in H^{g,t}$ *and all* $R_t \in \Upsilon$ *together with* $\sigma_{i,t}^{worst}(h^{i,t}, y_t^i) = 0$ *for all* $h^{i,t} \in H^{i,t}$*, all* $y_t^i \in [1, \overline{y}]$ *and all* $i \in I$ *is a Perfect Bayesian Equilibrium. It yields the lowest payoff amongst Perfect Bayesian equilibria.*

Proof: In the appendix.

Denote by V^{worst} the payoff generated by σ^{worst}.

Given the results in Propositions 5 and 6, in this section we analyze a class of government strategies that depends on both public and private histories. We restrict attention to government strategies that condition behavior on public histories and the most recent realization of the private shock, but not on the entire history of private shocks.[24] Let $\mathring{\Sigma}$ be the set of strategy profiles that conform with the restriction explained above. A strategy $\sigma \in \mathring{\Sigma}$ induces a stochastic outcome path. Given $\sigma \in \mathring{\Sigma}$, let $\tau^{t-1}(R^{t-1})$ be the t-period public history induced by σ_G and the sequence of shocks R^{t-1}.

Definition 10. *A sequence of functions $\mathcal{A} \equiv \{\tau_t, X_t, Y_t\}_{t=0}^{\infty}$ is the stochastic* ***aggregate outcome*** *induced by a strategy profile $\sigma \in \mathring{\Sigma}$ if:*

(i) $\tau_0(R_0) = \sigma_{G,0}(R_0)$, *and* $\tau_t(R^t) = \sigma_{G,t}(\tau^{t-1}(R^{t-1}), R_t)$;

(ii) $X_t(R^t) = \int_1^{\overline{y}} x_t^i(R^t, y_t^i) f(y_t^i|R_t) dy_t^i$ *and* $Y_t(R^t) = \int_1^{\overline{y}} y_t^i x_t^i(R^t, y_t^i) f(y_t^i|R_t) dy_t^i$, *where* $x_t^i(R^t, y_t^i)$ *is given by:*

$$x_t^i(R^t, y_t^i) = \begin{cases} \omega & \text{if } y_t^i E(1-\tau_t(R^t)|y_t^i) - 1 > 0 \\ [0,\omega] & \text{if } y_t^i E(1-\tau_t(R^t)|y_t^i) - 1 = 0 \\ 0 & \text{otherwise} \end{cases} \tag{28}$$

Notice that in Definition 10 the tax functions are potentially non-stochastic. Moreover, there is no restriction that constrains the tax functions to be time-stationary.

Proposition 8 gives a full characterization of aggregate allocations induced by profiles $\sigma \in \mathring{\Sigma}$. In order to state the proposition, given an stochastic **aggregate**

[24] In doing so we follow the same approach of both Sleet and Yeltekin (2006) and Athey, Atkeson, and Kehoe (2005).

outcome $\mathcal{A}$ of a profile $\sigma \in \mathring{\Sigma}$, let $V(\mathcal{A}) \equiv V(\sigma)$ denote the time zero expected payoff for the government. Similarly, denote by $V(\mathcal{A}|\tau^t(R^t))$ the continuation payoff from $\mathcal{A}$ after the public history $\tau^t(R^t)$. We also define, for a function $\tau_t \subseteq \mathcal{A}$, the set $\tau_t(\tau^{t-1}(R^{t-1}), \Upsilon) \equiv \left\{\hat{\tau}|\hat{\tau} = \tau_t(\tau^{t-1}(R^{t-1}), R) \text{ for some } R \in \Upsilon\right\}$ consisting of all values for taxes that can be assigned in period t through the mapping $\tau_t(\cdot)$. Finally, let $W^d(R_t, X(R^t), Y(R^t))$ the best deviation that the government can achieve after both the shock R_t and investment decisions $X(R^t)$ (together with aggregate output $Y(R^t)$) is realized. By Assumption 2, such deviation sets taxes to be equal 100%. Then we have the following:

Proposition 8. $\mathcal{A} \equiv \{\tau_t, X_t, Y_t\}_{t=0}^{\infty}$ *is the stochastic aggregate outcome of a Perfect Bayesian Equilibrium* $\sigma \in \mathring{\Sigma}$ *if and only if the following conditions are satisfied:*

(1) $\forall t$, $X_t(R^t) = \int_1^{\overline{y}} x_t^i(R^t, y_t^i) f(y_t^i|R_t) dy_t^i$ *and* $Y_t(R^t) = \int_1^{\overline{y}} y_t^i x_t^i(R^t, y_t^i) f(y_t^i|R_t) dy_t^i$, *where* $x_t^i(R^t, y_t^i)$ *is given by (28);*

(2) $\forall t, \tau^t(R^t)$, $V(\mathcal{A}|\tau^t(R^t)) \geq (1-\beta) W^d(R_t, X(R^t), Y(R^t)) + \beta V^{worst}$;

(3) $\forall t, \tau^{t-1}(R^{t-1}), R_t, \hat{\tau} \in \tau_t(\tau^{t-1}(R^{t-1}), \Upsilon)$:

$$\begin{aligned}(1-\beta)W(R_t, \tau_t(\tau^{t-1}(R^{t-1}), R_t), X(R^t), Y(R^t)) +& \\ \beta V(\mathcal{A}|\tau^{t-1}(R^{t-1}), \tau_t(\tau^{t-1}(R^{t-1}), R_t)) &\geq \\ (1-\beta)W(R_t, \hat{\tau}, X(R^t), Y(R^t)) + \beta V(\mathcal{A}|\tau^{t-1}(R^{t-1}), \hat{\tau})&\end{aligned}$$

Proof: In the appendix.

Condition (2) in the previous proposition comes from the standard reversion to the worst equilibrium in the case that the government deviates from a prescribed action along the path of play. Condition (3) is an incentive compatibility constraint that prevents the government to make profitable deviations along the path by misrepresenting its private information.

Let Λ be the set of aggregate allocations $\mathcal{A}$ that satisfies the conditions in proposition 8. For a given value of the discount factor β, let Ψ_β be the set of equilibrium payoffs of the repeated game that can be supported by some aggregate allocation $\mathcal{A}$:

$$\Psi_\beta = \{v^* : \exists \mathcal{A} \in \Lambda \text{ and } v^* = V(\mathcal{A})\}$$

The next lemma is a standard recursive result implied by the restriction of equilibria within the set $\mathring{\Sigma}$.

Lemma 11. *A payoff v^* is supported by the aggregate outcome $\mathcal{A}$ if and only if there exists functions $\{\tau, X, Y, v'\}$ with $\tau : \Upsilon \to \mathbb{R}_+$, $X : \Upsilon \to \mathbb{R}_+$, $Y : \Upsilon \to \mathbb{R}_+$ and $v' : \Upsilon \to \mathbb{R}_+$ such that:*

1. $v = \sum_{R \in \Upsilon} P(R)[(1-\beta)W(R, X(R), \tau(R), Y(R)) + \beta v'(R)]$

2. $\forall R \in \Upsilon,\ \ (1-\beta)W(R, X(R), \tau(R), Y(R)) + \beta v'(R) \geq (1-\beta)W^d(R, X(R), Y(R)) + \beta V^{worst}$

3. $\forall R, \hat{R} \in \Upsilon,\ \ (1-\beta)W(R, X(R), \tau(R), Y(R)) + \beta v'(R) \geq$
$$(1-\beta)W(R, X(R), \tau(\hat{R}), Y(R)) + \beta v'(\hat{R})$$

4. $\forall R \in \Upsilon,\ \ v'(R) \in \Psi_\beta$

5. $X(R) = \int_1^{\overline{y}} x^i(y^i) f(y^i|R) dy^i$ *and* $Y(R) = \int_1^{\overline{y}} y^i x^i(y^i) f(y^i|R) dy^i$, *where* $x^i(y^i)$ *is given by:*

$$x^i(y^i) = \begin{cases} \omega & \textit{if } y^i E(1-\tau(R)|y^i) - 1 > 0 \\ [0,\omega] & \textit{if } y^i E(1-\tau(R)|y^i) - 1 = 0 \\ 0 & \textit{otherwise} \end{cases}$$

For a given list of functions $\{\tau, X, Y, v'\}$ satisfying the conditions in lemma 11, let $\mathcal{OP}(\tau, R)$ be the aggregate allocations $(X(R), Y(R))$ generated by the optimal individual decisions according to condition (5) above. In order to save notation, we define $W(\tau(R), Z(\tau, R)) \equiv W(R, X(R), \tau(R), Y(R))$.

Next we restrict attention to the case in which the government's private shock can take only two possible values. From now on, τ and V will be vectors with each entry being a tax (or continuation value) contingent on R. Given both proposition 8 and lemma 11, the best equilibrium solves the following problem:

$$T = \max_{\{\tau, V\}} \sum_{s=L,H} P(R_s)[(1-\beta)W(\tau(R_s), Z(\tau, R_s)) + \beta V(\tau(R_s))] \tag{PR}$$

subject to:

for all $s = L, H$,

(IC-On) $\quad (1-\beta)W(\tau(R_s), Z(\tau, R_s)) + \beta V(\tau(R_s)) \geq (1-\beta)W(\tau(R_{-s}), Z(\tau, R_s)) + \beta V(\tau(R_{-s}))$

(IC-Off) $\quad (1-\beta)W(\tau(R_s), Z(\tau, R_s)) + \beta V(\tau(R_s)) \geq (1-\beta)W(1, Z(\tau, R_s)) + \beta V^{worst}$

(E) $\quad V(\tau(R_s)) \in \Psi_\beta$

(OP) $\quad Z(\tau, R_s) \in \mathcal{OP}(\tau, R_s)$

The constraint (IC-On) is the incentive compatibility constraint on the equilibrium path. That is, if the equilibrium strategy implies that on the equilibrium path two different taxes are used, and since R is virtually not observable, it has to be on the government's best interest to do it so. Constraint (IC-Off), or sustainability constraint, makes sure that the governments does not want to use a tax that is not contemplated on the equilibrium path. This is only possible if the payoff for any equilibrium tax is larger than the best deviation ($\tau = 1$ for any level of investment) plus the worst possible continuation value. The third constraint requires that every continuation value can be implemented as an equilibrium for some strategy pair and belief system. Finally, the last constraint imposes the optimality of households' decision rules.

Next we show that the above problem can be reduced to a more simple static maximization problem. Let $\bar{V} \equiv \sup\{\Psi_\beta\}$ and consider the following static problem,

$$\hat{T} = \max_{\{\tau,V\}} \sum_{s=L,H} P_{R_s}[(1-\beta)W(\tau(R_s), Z(\tau, R_s)) + \beta V(\tau(R_s))] \qquad \text{(PS)}$$

subject to: for all $s = L, H$,

(IC-ON-S) $\quad (1-\beta)W(\tau(R_s), Z(\tau, R_s)) + \beta V_s \geq (1-\beta)W(\tau(R_{-s}), Z(\tau, R_s)) + \beta V_{-s},$

(OP-S) $\quad Z(\tau, R_s) \in \mathcal{OP}(\tau, R_s)$ for all $\tau \in [0,1]^2$,

(E-S) $\quad V_s \in [V^{worst}, \bar{V}$

The next result states a key relation between problems PS and PR. In what follows, let $\hat{T}(\tau^*, V^*)$ be the value of the problem PS when its solution is given by (τ^*, V^*). $T(\cdot,\cdot)$ is defined in a similar fashion.

Proposition 9. *Let (τ^*, V^*) be a solution to problem PS and (τ^{**}, V^{**}) be the solution to PR. Then, there exists $\beta^* \in (0,1)$ such that for all $\beta \geq \beta^*$,*

i) $\hat{T}(\tau^*, V^*) \geq T(\tau, V)$ for all τ and V.

ii) If $V^* \in \Psi_\beta$ then $\hat{T}(\tau^*, V^*) = T(\tau^{**}, V^{**})$.

Proof: First notice there are two differences between PS and PR. First, the constraint (IC-Off), present in PR, is not included in PS. Second, the constraint (E) in PR requires that the chosen continuation values belong to the equilibrium value set, while PS only requires that the continuation values V_i belong to a convex and compact set. Notice that if β is large enough the constraint (IC-Off) would not be binding and we can disregard it. Thus, from this point of view, both problems are equivalent. But, by construction $\Psi_\beta \subseteq [V^{worst}, \bar{V}]$, therefore since the objective function is the same in both problems, and the feasible set in PR is smaller than in PS, part *i)* of the proposition follows. In addition, part *ii)* is immediate. If the solution to PS is feasible in PR then it must be the case that this value is the maximum ∎

We say that an strategy profile $\sigma \in \Sigma$ is **public** if for all t, $\tau_{t-1} \in [0,1]$, $h^{g,t} \in H^{g,t}$

and $R, \hat{R} \in \Upsilon$ we have that $\sigma_{G,t}(h^{g,t-1}, \tau_{t-1}, R) = \sigma_{G,t}(h^{g,t-1}, \tau_{t-1}, \hat{R})$. The main result of this section shows that, when agents discount the future high enough, the best Perfect Bayesian Equilibrium (within the class of strategies that we consider) is achieved through the use of public strategies by the government. As a byproduct, the best equilibria is inefficient.

Proposition 10. *If $\beta \geq \beta^*$ then, the solution to PS implies $\tau_L = \tau_H = \tau^B$ and $V_L = V_H = \bar{V}$. Moreover,*

1) $\tau^B = argmax_{\{\tau, Z(\tau,R_s) \in \mathcal{OP}(\tau,R_s)\}} \Big\{ E_R \big[W(\tau, Z(\tau, R_s)) \big] \Big\}$

2) $\bar{V} = E_{R_s} \big[W(\tau^B, Z(\tau^B, R_s)) \big]$

3) $\bar{V} \in \Psi_\beta$.

Proof: The lagrangian for this problem is

$$L = \sum_{s=L,H} P_s [W(\tau_s, Z(\tau, R_s)) + \beta V_s] + \sum_{s=L,H} \lambda_s [W(\tau_s, Z(\tau, R_s)) + \beta(V_s - V_{-s}) - W(\tau_{-s}, Z(\tau, R_s))] + \sum_{s=L,H} \gamma_s \beta [\bar{V} - V_s]$$

First notice that we are not considering the constraint that $V_s \geq V^{worst}$. This is guaranteed for a β large enough. Second, we do not need to consider the case in which both $\gamma_s > 0$. Because $W(.)$ is strictly increasing in τ, given interior aggregate allocations, the (IC-ON-S) would imply $\tau_L = \tau_H$. In the same way, if $\lambda_s > 0$ in both states, then both (IC-ON-S) would be binding, and therefore $\tau_L = \tau_H$. Thus, we only need to consider cases in which only one λ_s and only one γ_s can be strictly positive.

The first order conditions with respect to V_s are,

$$P_L + \lambda_L - \lambda_H - \gamma_L = 0$$

$$P_H + \lambda_H - \lambda_L - \gamma_H = 0$$

If $\gamma_L = \gamma_H = 0$ the equations above imply $P_L = -P_H$ which is not possible. That is, in a best equilibrium, in at least one state, the continuation value has to be a best equilibrium. Thus, we need to consider two cases.

Case 1: Suppose $\gamma_H > 0$ (hence $V_H = \bar{V}$ and $V_L \leq \bar{V}$), then the above equations imply $\lambda_H = P_L + \lambda_L$ or $\lambda_H > \lambda_L$, since at least one multiplier has to be zero, it follows that $\lambda_H > 0$. One can see by using (IC-ON-S) that in this case $\tau_L \geq \tau_H$ (the best equilibrium requires smaller continuation values for larger taxes).

Case 2: $\gamma_L > 0$ ($V_L = \bar{V}$, $V_H \leq \bar{V}$) and $\lambda_L > 0$. Using a similar argument it follows that $\tau_L \leq \tau_H$ in this case.

Consider case 1. Replacing the binding constraints in the objective function the problem becomes:

$$\hat{T} = \max_{\{\tau_L, \tau_H, V_L\}} P_L W(\tau_L, Z(\tau, R_L)) + P_H W(\tau_L, Z(\tau, R_H)) + \beta V_L$$

subject to;

$$W(\tau_H, Z(\tau, R_H)) + \beta \bar{V} \geq W(\tau_L, Z(\tau, R_H)) + \beta V_L$$

$$Z(\tau, R_s) \in \mathcal{OP}(\tau, R_s) \text{ for all } s$$

$$V_L \in [V^{worst}, \bar{V}]$$

But then again, either $V_L = \bar{V}$ or the (IC-ON-S) is binding, in both cases the solution implies $\tau_L = \tau_H$. A similar argument can be used to show that the second candidate solution implies the same result. Therefore, in any case $V_s = \bar{V}$. Then, maximizing the return function (imposing the additional constraint that taxes are equal) gives the first part of the proposition. The second part of the proposition follows from the fact that the maximum value is the summation of the period by period payoff, then if β is large enough this would be an equilibrium. ∎

Proposition 10 shows that any best equilibrium, within the constrained class of strategies that we analyze here, can be implemented with public strategies when agents discount the future high enough. Denote the best public equilibrium $\sigma^{BP} \in \Sigma$. Taxes then are a deterministic function of its own past history. Agents are able to predict perfectly the tax that they would have to pay after investing:

$$\sigma_{i,t}^{BP}(h^{i,t}, y^i) = \begin{cases} \omega & \text{if } y(1-\tau_t(\tau^{t-1})) - 1 > 0 \\ [0,\omega] & \text{if } y(1-\tau_t(\tau^{t-1})) - 1 = 0 \\ 0 & \text{otherwise} \end{cases}$$

The best equilibrium payoff under pure-public strategies for government is given by:

$$V^{BP} = V(\sigma^{BP}) = \max_{\tau_t(\tau^{t-1})} \sum_{t=0}^{\infty} \beta^t E[W(\tau_t, R_t| \left\{\sigma_{i,t}^{BP}\right\}_{i\in I})]$$

The best payoff under commitment, or the payoff of the Ramsey equilibrium, is given by:

$$V(\sigma^*) = \sum_{t=0}^{\infty} \beta^t E[W(\sigma^*(R_t), R_t| \left\{\sigma_{i,t}^{*}\right\}_{i\in I})]$$

Proposition 11. *There exists $\hat{\beta} < 1$ such that for all $\beta \in [\hat{\beta}, 1)$,$V^{BP} < V(\sigma^*)$.*

Proof: In the appendix.

The proposition follows because the Ramsey equilibrium exhibits taxes that are state dependent. Furthermore, the difference between the best one-period payoff under public equilibria and the Ramsey does not depend on the size of the discount factor.

19 About the Observability of Government Spending

The fact that neither the government spending nor the other aggregate variables are observable can be rationalized in the following way. Suppose that in stage 4 of the timing the government does not observe either the aggregate quantities or the average return on investment but a signal S of the true state of the economy. This signal can take on two values: $S \in \{S_L, S_H\}$ with $Pr(S_m|R_m) = 1 - \epsilon$ and $Pr(S_m|R_{-m}) = \epsilon$ for $\epsilon \in (0, 1)$ and $m = L, H$. After all players have made their decisions everyone observes the aggregate quantities. The new timing would be,

1. R is realized;
2. y^i is drawn for each i;
3. Each individual chooses x^i;
4. Government observes S and sets τ;
5. X and Y are observed by every player and the public good is produced;
6. G is observed and consumption takes place.

In the present environment, or an environment similar to Sleet and Yelketin (2006), the signal can be interpreted as an indicator of the actual state of the economy. Therefore, $1-\epsilon$ would be the precision of that indicator. In a monetary policy environment as Athey et. al. (2005) or Canzoneri (1985) S could be interpreted as the monetary authority forecast of future inflation. Under this interpretation, when $\epsilon = 0$ the indicator is perfectly informative and when $\epsilon = 1/2$ the indicator carries no information. The definition of a Ramsey equilibrium in this environment is equivalent to that in Section 4, but now the optimal tax policy will depend on S rather than on R. However, it is easy to see that this new Ramsey policy and the new Ramsey payoff would converge to that in Section 4 as ϵ goes to zero. Propositions 1 and 2 remain unchanged with R replaced by S in the statements. As before, for the equilibrium strategy to improve upon the payoff it must be a function of at least one publicly observed object. The main difference between this environment and the one developed in the paper is that now equilibrium strategies could depend on the past history of R. As long as R is *i.i.d.* over time, past realizations of the average productivity carry no information about the future behavior of R. Thus, conditioning on the past realizations of R generates no gains other than the possibility of coordination. Given that coordination was already possible in the original environment (when equilibrium strategies depend on publicly observed variables) the conclusions remains unaltered. That is, it is possible to show that Proposition 6 holds for every $\epsilon > 0$. However, if R were not *i.i.d.* over time it could be the case that the optimal policy without commitment is not invariant over time.

20 Conclusion

In this paper we show how small changes in the informational assumptions can have drastic consequences for both the set of equilibrium strategies and the set of equilibrium payoffs in a macro game without commitment. First we show that every equilibrium strategy has to depend on some information that is **publicly** known for every agent in the economy, otherwise no coordination is possible. In addition, when we analyze equilibria that depend on both public and private histories we found that in the best equilibrium the government **does not use** its private information. As a result, for any discount factor strictly smaller than one, the payoff of the best equilibrium without commitment is always strictly smaller than the payoff with commitment. Moreover, this distance does not approach zero as the discount factor approaches one. In this sense, the welfare in a economy without commitment is uniformly bounded away from the efficient one.

The results of this paper support the arguments for strong institutions that tie the hands of policy makers. To endowed governments with full discretion and to impose the right incentives to avoid deviations from optimal policies could be impossible or too costly. In this sense, the original recommendation of Kydland and Prescott (1977) is still valid.

The implication of this paper apparently contradicts the fact that most policies react to the state of the economy. However, according to the interpretation of Section 19 this will not constitute a contradiction. It could be optimal for the government to react to **past** (and publicly known) states of the economy as long as it provides information about future states. What the policy maker looses is the possibility of fine tuning using **not perfectly precise** signals. On the other hand, it is fairly common to find examples of policy makers that have been institutionally banned from the use of discretion, e.g, the implementation of a currency board system. This usually happens when the policy maker has a "bad reputation", like Argentina in the 90's. We think that future research on time inconsistency problems should include the possibility for the policy maker of building reputation. This line of research, if successful, will provide a more precise answer to this kind of problems.

References

Alesina, A., F. Campante, and G. Tabellini (2008): "Why is Fiscal Policy Often Procyclical?," *Journal of the European Economic Association*, 6(5), 1006–1036.

Alesina, A., and D. Rodrik (1994): "Distributive Politics and Economic Growth," *The Quarterly Journal of Economics*, 109, 465–490.

Athey, S., A. Atkeson, and P. Kehoe (2005): "The Optimal Degree of Discretion in Monetary Policy," *Econometrica*, 73(5), 1431–1475.

Azzimonti, M., E. de Francisco, and P. Krusell (2006): "Median-voter Equilibria in the Neoclassical Growth Model under Aggregation," *Scandinavian Journal of Economics*, 108(4), 587–606.

——— (2008): "Aggregation and Aggregation," *Journal of the European Economic Association*, 6(2–3), 381–394.

Bassetto, M., and J. Benhabib (2006): "Redistribution, Taxes, and the Median voter," *Review of Economic Dynamics*, (9).

Benhabib, J., and A. Przeworski (2006): "The Political Economy of Redistribution under Democracy," *Economic Theory.*

Bernheim, B. D., and S. N. Slavov (2008): "A Solution Concept for Majority Rule in Dynamic Settings," *The Review of Economic Studies, Forthcoming.*

Chamley, C. (1986): "Optimal Taxation of Capital Income in General Equilibrium with Infinite Lives," *Econometrica*, 54, 607–622.

Chari, V. V., L. Christiano, and P. Kehoe (1994): "Optimality fiscal policy in a business cycle model," *Journal of Political Economy*, 102, 617–652.

Chari, V. V., and P. Kehoe (1993): "Sustainable Plans and Mutual Default," *Review of Economic Studies*, 60(1), 175–95.

Chari, V. V., and P. J. Kehoe (1990): "Sustainable Plans," *Journal of Political Economy*, 98(4), 783–802.

——— (1999): "Optimal Fiscal and Monetary Policy," *in J.B. Taylor and M. Woodford (ed.), Handbook of Macroeconomics.*

Corbae, D., P. D'Erasmo, and B. Kuruscu (2008): "Politico-economic consequences of rising wage inequality," *Journal of Monetary Economics, forthcoming.*

Eckstein, Z., and E. Nagypal (2004): "The Evolution of US Earnings Inequality: 1961-2002," *Federal Reserve Bank of Minneapolis Quarterly Review*, 28, 10–29.

Kramer, G. (1973): "On a Class of Equilibirum Conditions for Majority Rule," *Econometrica*, 41(2).

Krusell, P., and J. V. Rios-Rull (1999): "On the Size of U.S. Government: Political Economy in the Neoclassical Growth Model," *The American Economic Review*, 89, 1156–1181.

KYDLAND, F., AND E. PRESCOTT (1977): "Rules Rather Than Discretion: The Inconsistency of Optimal Plans," *Journal of Political Economy*, 85(3), 473–91.

MAILATH, G. J., AND L. SAMUELSON (2006): *Repeated Games and Reputations. Long-run relationships.* Oxford University Press.

MCDANIEL, C. (2007): "Average tax rates on consumption, investment, labor and capital in the OECD 1950-2003," *Working Paper*.

MELTZER, A., AND S. RICHARD (1981): "A Rational Theory of the Size Government," *Journal of Political Economy*, 89, 914–927.

PERSSON, T., AND G. TABELLINI (1994): "Is Inequality Harmful for Growth?," *American Economic Review*, 84, 600–621.

PHELAN, C., AND E. STACCHETTI (2001): "Sequential Equilibria in a Ramsey Tax Model," *Econometrica*, 69(6), 1491–1518.

PIGUILLEM, F., AND A. L. SCHNEIDER (2007): "Optimal Fiscal Policy and the Extensive Margin of Labor Supply," *(www.econ.umn.edu/ ~ als/Research/Taxescornersol.pdf)*.

SLEET, C. (2001): "On Credible Monetary Policy and Private Government Information," *Journal of Economic Theory*, 99, 338–376.

SLEET, C., AND S. YELTEKIN (2006): "Optimal taxation with endogenously incomplete debt markets," *Journal of Economic Theory*, 127(1), 36–73.

Werning, I. (2007): "Optimal Fiscal Policy with Redistirbution," *Quarterly Journal of Economics*, 122.

Part V

Appendix

21 Appendix of Chapter 1

21.1 Proof of Lemma 1:

For the necessity part of the Lemma, it remains to show that the individual shares integrate to one regardless of the normalization of the initial price p_0. From (4)-(7), let $\omega_c(\theta)$ be the individual share of type θ on the aggregate consumption:

$$\begin{aligned} c(s^t;\theta) &= \omega_c(\theta)C(s^t) \\ \theta - l(s^t;\theta)/\theta &= \omega_c(\theta)[1-L(s^t)] \end{aligned}$$

Using both (4)-(7) and the above representation of individual allocations in the budget constraint yield:

$$\omega_c(\theta) = (1-\beta)\left[\widetilde{W}_0(\theta,T,\tau_{k0}) + \theta\sum_{t,s^t}\beta^t \Pr(s^t)\left[\frac{(1-\alpha)}{(1-L(s^t))}\right]\right]$$

In what follows below, let $F_l(s^t)$ and $F_k(s^t)$ be the marginal product of labor and capital respectively. The intratemporal optimality condition for each individual can be expressed as:

$$(1-L(s^t))(1-\tau_l(s^t))F_l(s^t) = \frac{1-\alpha}{\alpha}C(s^t)$$

Then:

$$\begin{aligned}
C(s^t) &= \alpha[(C(s^t) - F_l(s^t)L(s^t)) + \tau_l(s^t)F_l(s^t)L(s^t) + (1-\tau_l(s^t))F_l(s^t)] \\
\mathrm{Pr}(\mathrm{s}^t)\beta^t &= \mathrm{Pr}(\mathrm{s}^t)\beta^t \frac{\alpha}{C(s^t)}[-K(s^t) + (1 + (1-\tau_k(s^t))(F_k(s^t) - \delta))K(s^{t-1}) + \hat{T}_t + (1-\tau_{lt})F_l(s^t)]
\end{aligned}$$

where $\hat{T}_t \equiv \tau_k(s^t)F_k(s^t)K(s^{t-1}) + \tau_l(s^t)F_l(s^t)L(s^t)$.

Since the final expression above is true in each period, using the intertemporal optimality conditions we have:

$$\frac{1}{1-\beta} = \frac{\alpha}{C_0}(1 + (1-\tau_{k0})(F_{k0} - \delta))K_0 - T + \sum_{t,s^t} \mathrm{Pr}(\mathrm{s}^t)\beta^t \left[\frac{1-\alpha}{1-L(s^t)}\right]$$

or

$$1 = (1-\beta)\left[\frac{\alpha}{C_0}(1 + (1-\tau_{k0})(F_{k0} - \delta))K_0 - T + \sum_{t,s^t} \mathrm{Pr}(\mathrm{s}^t)\beta^t \left(\frac{1-\alpha}{1-L(s^t)}\right)\right]$$

But this is equivalent to $\int_\theta \omega_c(\theta) f(\theta) d\theta = 1$.

Next we prove the sufficiency of conditions (1)-(3). First, use the function $\varphi(\cdot)$ and Z to construct the individual consumption and labor allocations. Set after-tax prices as:

$$p(s^t) = \beta^t \mathrm{Pr}(s^t) U_c^{CE}(s^t) = \beta^t \mathrm{Pr}(s^t) E(\varphi)\alpha / C(s^t), \;\; p(s_0) = \alpha / C_0$$

$$p(s^t)w(s^t)(1-\tau_l(s^t)) = \beta^t \mathrm{Pr}(s^t) E(\varphi)(1-\alpha)/(1-L(s^t))$$

And

$$r(s^t) = F_k(s^t), w(s^t) = F_L(s^t), p(s^t) = \sum_{s_{t+1}} p(s^{t+1})R(s^{t+1})$$

If $0 < \frac{\varphi(\theta)}{E(\varphi)} \leq \frac{\theta}{1-L(s^t)}$ for all s^t, then one can check, using the solution to the static problem, that the following necessary first order conditions are met for all s^t and $l \in [0,1]$:

$$\left[U_l(c^*(s^t;\theta),1-l^*(s^t;\theta)/\theta)+U_c(c^*(s^t;\theta),1-l^*(s^t;\theta)/\theta)\left(w(s^t)(1-\tau_l(s^t))\right)\right][l-l^*(s^t;\theta)]\leq 0$$

The transversality condition (Tvc) $\lim_{t\rightarrow\infty}\sum_{s^t} p(s^t)k(s^t;\theta)=0$ is satisfied because it can be shown that individual capital allocations are an affine function of the aggregate capital stock. At the equilibrium prices, the Tvc is met, since the aggregate allocations are bounded in the product topology. Finally, using 11, we can get the budget constraint back. Condition (2) in the competitive equilibrium definition is satisfied by construction. As usual, the government budget constraint can be recovered using a version of the Walras' law. Taxes on capital can be constructed in many ways, and taxes on labor are constructed using the definition of prices and $w(s^t)=F_L(s^t)$ ■

21.2 Proof of Lemma 2:

Let $\tilde{\tilde{\Xi}}$ be the set of allocations $Z\equiv(\{C(s^t),L(s^t),K(s^t)\}_{t\geq 0},T\leq 0,\tau_{k0})$ with aggregate labor allocation bounded away from zero, the resources constraint satisfied for all periods, and the Euler equation satisfied with weak inequality.

For any $Z\in\tilde{\tilde{\Xi}}$, let $\underline{L}(Z)\equiv\inf\{L(s^t)\}_{t\geq 0}$ and define $\underline{\theta}(Z)$ to be the solution to:

$$\inf \quad \theta \quad \text{s.t.} \begin{cases} \theta\in[0,1] \\ \varphi(Z;\theta)<\frac{1}{1-\underline{L}(Z)} \end{cases}$$

Claim: $\underline{\theta}(Z)$ is bounded away from 1 for all $Z\in\tilde{\tilde{\Xi}}$.

Proof of the claim: Because of the linearity of $\varphi(\cdot)$ in types, it follows that $\varphi(Z;\theta = 1) = 1$. $\{L(s^t)\}_{t\geq 0}$ is bounded away from zero, and therefore $\frac{1}{1-\underline{L}(Z)} \geq \frac{1}{1-\epsilon}$ for some $\epsilon > 0$. The claim follows.

Define $\widehat{\theta} \equiv \sup \{\underline{\theta}(Z) : Z \in \widetilde{\Xi}\}$. Because of the claim above, $\widehat{\theta} < 1$. Then it is straightforward to check that $\widehat{\theta}$ has the property stated in the Lemma. In particular, if $\underline{\theta}(Z)$ satisfies the second constraint in the inf problem above, then it satisfies that constraint for all $L(s^t) \subseteq Z)$ ■

21.3 Proof of Lemma 3:

If the statement is not true, then in the solution to $P(M)$ we have $T^* > 0$. The value of the program $P(M)$ can be written as:

$$P(M) = \frac{1}{1-\beta}\log(\varphi(Z^*, \theta^m)) + V(Z^*)$$

where $\varphi(Z^*;\theta^m)$ and $V(Z^*)$ are given by (11) and the last part of (13) respectively, evaluated at Z^*. Now fix $\hat{T} = 0$. For any $\hat{Z} \in \Xi$ with $\hat{T} = 0$

$$\begin{aligned}
\varphi(Z;\theta^m) &= (1-\beta)\theta^m \left[\frac{\alpha}{\hat{C}_0}\hat{R}_0\frac{(\gamma_k)}{\theta^m} + \frac{\alpha}{\hat{C}_0}\hat{R}_0(K_{-1}-\gamma_k) + \widehat{UL}\right] \\
&= \theta^m + \varepsilon(\hat{Z})
\end{aligned}$$

for some $\varepsilon(\hat{Z}) > 0$. Both the last equality and the fact that $\varepsilon(\hat{Z}) > 0$ come from $E(\varphi(\theta)) = 1$, $\theta^m < 1$ and $\gamma_k > 0$. Next, define the feasible allocation $\hat{Z} \in \Xi$ with $\hat{T} = 0$ as:

$$\hat{Z} \in \text{argmax}_{\{C,L,K,\tau_{k0}\}} V(Z)$$

$$
\text{s.t.} \begin{cases} C(s^t) + K(s^t) \leq F\Big(L(s^t), K(s^{t-1}), s^t\Big) + (1-\delta)K(s^{t-1}) & \forall s^t \quad (\mathbf{RC}); \\ \frac{1}{C(s^t)} \geq \beta \sum_{s_{t+1}} \Pr(s_{t+1}|s^t) \frac{[1+(1-\overline{\tau})(F_k(s^{t+1})-\delta)]}{C(s^{t+1})} & \forall s^t \quad (\mathbf{UB}); \\ (1-\beta)\left[\frac{\alpha}{C_0} R_0 k_{-1}(\underline{\theta}) + \underline{\theta} UL\right] \geq 0 & (\mathbf{NN}); \\ \tau_{k0} \leq \overline{\tau}, \;\; K_{-1} \text{ given} \end{cases}
$$

Clearly, the constraint NN will never bind. Therefore the value of $\hat{Z}$ in terms of utility is given by:

$$
\begin{aligned}
\hat{P}(Z) &= \frac{1}{1-\beta} \log(\theta^m + \varepsilon(\hat{Z})) + \left\{ \max_{\{C,L,K,\tau_{k0}\} \in (RC, UB, \tau_{k0} \leq \underline{\tau})} V(Z) \right\} \\
&= \frac{1}{1-\beta} \log(\theta^m) + V(\hat{Z})
\end{aligned}
$$

Then, since Z^* solves $P(M)$, we have that:

$$
\begin{aligned}
P(\widehat{Z}) - P(M) = \frac{1}{1-\beta} \Big[& \log(\theta^m + \varepsilon(\hat{Z})) \\
& - \log\left((1-\beta)\left(\frac{\alpha}{C_0^*}(R_0^*(\gamma_k + (K_{-1} - \gamma_k)\theta^m)) - T^* + \theta^m UL^*\right)\right)\Big] + \\
& + V(\hat{Z}) - V(Z^*) \leq 0
\end{aligned}
$$

By definition of $\hat{Z}$ it must be the case that $V(\hat{Z}) - V(Z^*) \geq 0$. In addition notice that $\theta^m > (1-\beta)\left(\frac{\alpha}{C_0^*}(R_0^*(\gamma_k + (K_{-1} - \gamma_k)\theta^m)) - T^* + \theta^m UL^*\right)$. If not, we would have

$$
\theta^m \leq (1-\beta)(\frac{\alpha}{C_0^*}(R_0^*(\gamma_k + (K_{-1} - \gamma_k)\theta^m)) - T^* + \theta^m UL^*)
$$

or

$$
\theta^m \left[1 - (1-\beta)UL^* - (1-\beta)\frac{\alpha}{C_0^*} R_0^*(K_{-1} - \gamma_k)\right] \leq (1-\beta)\left[-T^* + \frac{\alpha}{C_0^*}\left(R_0^* \gamma_k\right)\right]
$$

Since the individual shares integrate to the unity (see proof of Lemma 1), it follows that $(1-\beta)(\frac{\alpha}{C_0^*}(R_0^*(\gamma_k + (K_{-1} - \gamma_k)) - T^* + UL^*) = 1$. Replacing this condition in the inequality above yields:

$$\theta^m(1-\beta)(-T^*) \leq (1-\beta)(-T^*)$$

But since $T^* > 0$ the above inequality implies $\theta^m \geq 1$, a contradiction. Therefore, $\theta^m > (1-\beta)\left(\frac{\alpha}{C_0^*}(R_0^*(\gamma_k + (K_{-1} - \gamma_k)\theta^m) - T^*) + \theta^m UL^*\right)$. This last strict inequality implies $P(\widehat{Z}) - P(M) > 0$, a contradiction ∎

21.4 Proof of Lemma 4:

The following slightly modifies the proof in Bassetto and Benhabib (2006).

If the claim is not true, then $\{C^*(s^t), K^*(s^t)\}_{s^t > s^{\tilde{t}}}$ does not satisfy the first order conditions in the following problem:

$$\max_{\{C(s^t),K(s^t)\}_{s^t>s^{\tilde{t}}}} \sum_{s^t > s^{\tilde{t}}} \beta^t \Pr(s^t)\left[\alpha \log(C(s^t)) + (1-\alpha)\log(1-L(s^t))\right]$$

$$\text{s.t.} \begin{cases} C(s^t) + K(s^t) + g(s^t) \leq F\Big(L(s^t), K(s^{t-1}), s^t\Big) + (1-\delta)K(s^{t-1}) \;\text{ for all }\; s^t > s^{\tilde{t}} \\ K^*(s^{\tilde{t}}),\, K_{-1},\, T^*,\, \tau_{k0} \text{ and } \{L^*(s^t)\}_{s^t > s^{\tilde{t}}} \text{ given} \end{cases}$$

Then it follows there must be an alternative allocation $\{\hat{C}(s^t), \hat{K}(s^t)\}_{s^t > s^{\tilde{t}}}$ satisfying the constraints above that yields a higher value for the return function.

Let $s^t \not> s^{\tilde{t}}$ with $t > \tilde{t}$ denote a history s^t that does **not** follow the history $s^{\tilde{t}}$.

Since the utility for the median type is increasing in the value of the utility for the mean type, it follows that $\{\hat{C}(s^t), \hat{K}(s^t)\}_{s^t > s^{\tilde{t}}}$ and K_{-1}, T^*, τ_{k0}, $\{L^*(s^t)\}_{t \geq 0}$,

$\{C^*(s^t), K^*(s^t)\}_{t<\tilde{t}}$ and $\{C^*(s^t), K^*(s^t)\}_{s^t \not\succ s^{\tilde{t}}}$ is a feasible allocation for the median voter's problem that improves the objective function, a contradiction ∎

21.5 Proof of Lemma 5:

By Lemma 4 constraint (NN) can be disregarded. Let μ be the multiplier associated with the constraint on τ_{k0}. Then the Foc for τ_{k0} generates:

$$\frac{(1-\theta^m)}{\varphi(\theta^m)} = \mu \frac{C_0}{\alpha(K_{-1}-\gamma_k)} \frac{1}{(F_{k_0}-\delta)} \tag{29}$$

Therefore $\mu > 0$, which implies that there is a corner solution for τ_{k0}. Next, define $R_0 = 1 + (1-\bar{\tau})(F_{k_0}-\delta)$. The first order conditions without considering the conditions in (UB) imply:

$$\frac{1}{C_0^2} R_0(K_{-1}-\gamma_k) \frac{(1-\theta^m)}{\varphi(\theta^m)} + \frac{1}{C_0} = \beta E \left[\frac{1-\delta+F_k(s^1)}{C(s^1)} \mid s_0 \right]$$

Constraint (UB) at s_0 is satisfied when:

$$\frac{1}{C_0^2} R_0(K_{-1}-\gamma_k) \frac{(1-\theta^m)}{\varphi(\theta^m)} - E\left[\frac{\bar{\tau}(F_k(s^1)-\delta)}{C(s^1)} \mid s_0 \right] \leq 0 \tag{30}$$

Because of the log utility function on consumption, any solution will have $C_0^* > 0$ regardless the size of $K_{-1}-\gamma_k \geq 0$. From Lemma 3 we have that $\theta^m < \varphi(Z^*, \theta^m) < 1$. Therefore the above holds when $\delta = 0$ and $K_{-1}-\gamma_k$ is sufficiently small.

As it is well known from Chari and Kehoe (1999), the process for taxes on capital income as a function of implementable allocations is not uniquely determined. In particular, as in Werning (2007), one such process can be constructed by:

$$\frac{1+(1-\tau_k(s^1))(F_k(s^1)-\delta)}{1-\delta+F_k(s^1)} = \frac{U_c(s_0)}{V_{c0}(\theta^m, Z)} \frac{V_{c1}(\theta^m, Z)}{U_c(s^1)} \text{ for all } s^1$$

where $V(\cdot)$ stands for the objective function in the median voter's problem. Then using (29) the expression yields:

$$\frac{1+(1-\tau_k(s^1))(F_k(s^1)-\delta)}{1-\delta+F_k(s^1)} = \frac{1/C_0}{\left[\mu\frac{R_0}{\alpha C_0(F_{k_0}-\delta)}\right]+1/C_0} \text{ for all } s^1$$

Clearly, the above equation implies that $1+(1-\tau_k(s^1))(F_k(s^1)-\delta) < 1-\delta+F_k(s^1)$, and in turn that $\tau_k(s^1) > 0$ for all s^1. This proves the second line.

Finally, the Foc's without considering UB satisfy the constraint for all $t \geq 2$. Furthermore, the implied taxes on capital returns are zero ▮

21.6 Median Voter Result and Characterization in the General Case

As in the main part of the paper, set $\varphi(\theta) \equiv 1/\lambda(\theta)$, where $\lambda(\theta)$ is the multiplier related to the present value budget constraint of type θ. Working with the first order conditions with respect to individual consumption and labor yields:

$$c(s^t,\theta) = \frac{\theta^{\frac{-(1-\alpha)(1-\sigma)}{\sigma}}\varphi^{1/\sigma}(\theta)}{\int_\Theta \theta^{\frac{-(1-\alpha)(1-\sigma)}{\sigma}}\varphi^{1/\sigma}(\theta)f(\theta)d\theta}C(s^t) = \omega_c(\theta)C(s^t)$$

$$1-\frac{l(s^t,\theta)}{\theta} = \frac{\theta^{\frac{\alpha(1-\sigma)-1}{\sigma}}\varphi^{1/\sigma}(\theta)}{\int_\Theta \theta^{\frac{-(1-\alpha)(1-\sigma)}{\sigma}}\varphi^{1/\sigma}(\theta)f(\theta)d\theta}[1-L(s^t)] = \omega_L(\theta)[1-L(s^t)]$$

and

$$\begin{aligned} p(s^t) &= \alpha\Phi^\sigma[C(s^t)^\alpha(1-L(s^t))^{1-\alpha}]^{-\sigma}\left[\frac{C(s^t)}{1-L(s^t)}\right]^{\alpha-1} \\ p(s^t)w(s^t)(1-\tau_l(s^t)) &= (1-\alpha)\Phi^\sigma[C(s^t)^\alpha(1-L(s^t))^{1-\alpha}]^{-\sigma}\left[\frac{C(s^t)}{1-L(s^t)}\right]^{\alpha} \end{aligned}$$

where $\Phi = \int_\Theta \theta^{\frac{-(1-\alpha)(1-\sigma)}{\sigma}}\varphi^{1/\sigma}(\theta)f(\theta)d\theta$.

Let $U(C(s^t), L(s^t)) \equiv \frac{\left[C(s^t)^\alpha(1-L(s^t))^{1-\alpha}\right]^{1-\sigma}}{1-\sigma}$. As in the logarithmic case, replacing prices and individual allocations in the budget constraint for each agent θ yields:

$$\sum_t \beta^t \text{Pr}(s^t)\omega_c(\theta)\alpha(1-\sigma)U(C(s^t), L(s^t) = U_{c0}\widetilde{W}_0(\theta, T, \tau_{k0})$$
$$+\sum_t \beta^t \text{Pr}(s^t)(1-\alpha)\left[C(s^t)^\alpha(1-L(s^t))^{1-\alpha}\right]^{-\sigma}\left(\frac{C(s^t)}{(1-L(s^t))}\right)^\alpha \theta\left[1-\omega_L(\theta)\left(1-L(s^t)\right)\right]$$

Let $UL = \sum_t \beta^t \text{Pr}(s^t)\left(C(s^t)\alpha(1-L(s^t))^{1-\alpha}\right)^{1-\sigma}\frac{(1-\alpha)}{1-L(s^t)}$. Then we can use the fact that $\omega_c(\theta) = \theta\omega_L(\theta)$ to get:

$$(1-\sigma)\omega_c(\theta)\sum_t \beta^t Pr(s^t)U(C(s^t), L(s^t)) = U_{c0}\widetilde{W}_0(\theta, T, \tau_{k0}) + \theta UL$$

or

$$(1-\sigma)\frac{\varphi(\theta)^{1/\sigma}}{\Phi}\theta^{\frac{-(1-\alpha)(1-\sigma)}{\sigma}}V(Z) = U_{c0}\widetilde{W}_0(\theta, T, \tau_{k0}) + \theta UL$$

where $V(Z) = \sum_t \beta^t Pr(s^t)U(C(s^t), L(s^t))$.

Therefore we have the following:

$$\omega_c(\theta) = \frac{[U_{c0}\widetilde{W}_0(\theta, T, \tau_{k0}) + \theta UL]}{(1-\sigma)V(Z)} \tag{31}$$

Since $\int_\Theta \omega_c(\theta)f(\theta)d\theta = 1$, the utility of household type θ in a particular competitive equilibrium can be written as:

$$V(Z, \theta^m) = \frac{[\chi(Z, \theta^m)]^{1-\sigma}}{\theta^{(1-\alpha)(1-\sigma)}}V(Z) \tag{32}$$

where

$$\chi(Z, \theta^m) \equiv \frac{U_{c0}[\widetilde{W}_0(\theta^m, T, \tau_{k0}) - E[\widetilde{W}_0(\theta, T, \tau_{k0})]] + (\theta^m - 1)UL + (1-\sigma)V(Z)}{(1-\sigma)V(Z)}$$

Given two competitive equilibrium allocations $Z, \hat{Z} \in \Upsilon$, type θ prefers Z to $\hat{Z}$ iff $V(Z;\theta) \geq V(\hat{Z};\theta)$, or alternatively, $\log(V(Z;\theta)/V(\hat{Z};\theta)) \geq 0$.

Equation (32) can be used to compute the ratio $V(Z;\theta)/V(\hat{Z},\theta)$ as

$$\frac{V(Z;\theta)}{V(\hat{Z};\theta)} = \left(\frac{[U_{c0}W_0(\theta,T,\tau_{k0}) + \theta UL]}{[\widetilde{U}_{c0}W_0(\theta,\hat{T},\hat{\tau}_{k0}) + \theta \hat{UL}]}\right)^{1-\sigma} \left(\frac{\hat{\Phi}V(Z)}{\Phi V(\hat{Z})}\right)^{\sigma}$$

Proposition 1.1: **(MVT - Inequality in both labor skills and initial wealth)** Suppose the initial wealth is an affine function of skills, i.e., $k_0(\theta) = \nu_1 + \nu_2\theta$. Consider any $Z, \hat{Z} \in \Xi$. If $\theta^m \in S_{Z,\hat{Z}}$, then either $[\underline{\theta}, \theta^m] \subseteq S_{Z,\hat{Z}}$ or $[\theta^m, \overline{\theta}] \subseteq S_{Z,\hat{Z}}$.

Proof. $W_0(\hat{T}, \hat{\tau}_{k0})$ can be written as $W_0(\hat{T}, \hat{\tau}_{k0}) = a + R\nu_2\theta$. Then consider the following derivative

$$\begin{aligned}
\frac{\partial \log\left(\frac{V(Z;\theta)}{V(\hat{Z};\theta)}\right)}{\partial\theta} &= (1-\sigma)\left[\frac{UL + \nu_2 R U_{c0}}{[U_{c0}W_0(T,\tau_{k0}) + \theta UL]} - \frac{\hat{UL} + \hat{R}\nu_2\hat{U}_{c0}}{[\hat{U}_{c0}W_0(\hat{T},\hat{\tau}_{k0}) + \theta\hat{UL}]}\right] \\
&= (1-\sigma)\left[\frac{\hat{a}\hat{U}_{c0}(UL + R\nu_2 U_{c0}) - aU_{c0}(\hat{UL} + \hat{R}\nu_2\hat{U}_{c0})}{[U_{c0}W_0(T,\tau_{k0}) + \theta UL][\hat{U}_{c0}W_0(\hat{T},\hat{\tau}_{k0}) + \theta\hat{UL}]}\right]
\end{aligned}$$

Therefore, as in the log case, the sign of the derivative does not depend on θ ∎

21.6.1 Characterization

The objective function for the median voter problem is given by:

$$V(Z,\theta^m) = V(Z)\left\{\frac{u_{c0}[\widetilde{W}_0(\theta^m,T,\tau_{k0}) - E[\widetilde{W}_0(\theta,T,\tau_{k0})]] + (\theta^m - 1)UL + (1-\sigma)V(Z)}{[(1-\sigma)V(Z)]}\right\}^{1-\sigma} \tag{33}$$

where $V(Z) = \sum_t \beta^t Pr(s^t) U^{CE}(C(s^t), L(s^t))$ and $UL = \sum_t \beta^t \Pr(s^t) u\Big(C(s^t), 1 - L(s^t)\Big)\frac{(1-\alpha)(1-\sigma)}{1-L(s^t)}$.

In the general case, problem P(M) becomes:

$$\max_{\{C,L,K,T,\tau_0\}} V(Z,\theta^m)$$

$$
\text{s.t.} \begin{cases} C(s^t) + K(s^t) + g(s^t) \leq F\Big(L(s^t), K(s^{t-1}), s^t\Big) + (1-\delta)K(s^{t-1}) \quad (\mathbf{RC}); \\ U_c(s^t) \geq \beta \sum_{s_{t+1}} \Pr(s_{t+1}|s^t) U_c(s^{t+1})[1 + (1-\overline{\tau})(F_k(s^{t+1}) - \delta)] \qquad (\mathbf{UB}) \\ U_{c0}^{CE}[\widetilde{W}_0(\underline{\theta}, T, \tau_{k0}) - E[\widetilde{W}_0(\theta, T, \tau_{k0})]] + (\underline{\theta} - 1)UL + (1-\sigma)V(Z) \geq 0 \quad (\mathbf{NN}) \\ \tau_{k0} \leq \overline{\tau} \end{cases}
$$

We use the notation of equation (32) to express the return function in the problem above as $V(Z, \theta^m) = [\chi(Z, \theta^m)]^{1-\sigma} V(Z)$.

It can be shown that the partial derivatives are given by:

$$
\hat{V}_C(s^t) = \beta^t \Pr(s^t) [\chi(Z, \theta^m)]^{1-\sigma} \left\{ \frac{(1-\sigma)}{\chi(Z, \theta^m)} \left[\frac{(\theta^m - 1)(1-\alpha)}{1 - L(s^t)} + 1 \right] + \sigma \right\} U_c(s^t) \tag{34}
$$

$$
\hat{V}_L(s^t) = \beta^t \Pr(s^t) [\chi(Z, \theta^m)]^{1-\sigma} \left\{ \frac{(\theta^m - 1)[(1-\alpha)(1-\sigma) - 1]}{[1 - L(s^t)]\chi(Z, \theta^m)} + \frac{1-\sigma}{\chi(Z, \theta^m)} + \sigma \right\} U_L(s^t) \tag{35}
$$

With some abuse of notation, let:

$$
a(s^t) = [\chi(Z, \theta^m)]^{1-\sigma} \left\{ \frac{(1-\sigma)}{\chi(Z, \theta^m)} \left[\frac{(\theta^m - 1)(1-\alpha)}{1 - L(s^t)} + 1 \right] + \sigma \right\}, \quad b(\theta^m) = \frac{-(\theta^m - 1)}{(1 - L(s^t))\chi(Z, \theta^m)}
$$

Lemma 9: If $\theta^m < 1$ then in any solution to P(M) we have $\theta^m \leq \chi(Z, \theta^m) < 1$.

Proof. Clearly $\chi(Z, \theta^m) < 1$ when $\theta^m < 1$, so we only need to show that $\theta^m < \chi(Z, \theta^m)$. It follows that $W_0(\theta, T, \tau_{k0}) = \gamma_k + (K_0 - \gamma_k)\theta - T$ because $R_0 = 1$ at the optimum. As in the proof of Lemma 4, in the solution to median voter problem (with $\theta^m < 1$) we have $T \leq 0$. Then it must be true that:

$$\begin{aligned}
0 &\leq U_{c0}\left[-(\theta^m-1)\gamma_k+(\theta^m-1)T\right] \\
&= U_{c0}\left[(\theta^m-1)(K_0-\gamma_k)-(\theta^m-1)K_0+(\theta^m-1)T\right] \\
&= U_{c0}\left[W_0(\theta^m,T,\tau_{k0})-E[W_0(\theta,T,\tau_{k0})]\right]-(\theta^m-1)U_{c0}E[W_0(\theta,T,\tau_{k0})] \\
&= U_{c0}\left[W_0(\theta^m,T,\tau_{k0})-E[W_0(\theta,T,\tau_{k0})]\right]-(\theta^m-1)(1-\sigma)V+(\theta^m-1)UL
\end{aligned}$$

Where $(1-\sigma)V = U_{c0}E[W_0(\theta,T,\tau_{k0})]+UL$ from the **MKT** constraint. The last inequality can be written as

$$U_{c0}\left[W_0(\theta^m,T,\tau_{k0})-E[W_0(\theta,T,\tau_{k0})]\right]+(\theta^m-1)UL \geq (\theta^m-1)(1-\sigma)V$$

or,

$$\frac{U_{c0}\left[W_0(\theta^m,T,\tau_{k0})-E[W_0(\theta,T,\tau_{k0})]\right]+(\theta^m-1)UL}{(1-\sigma)V}+1 \geq \theta^m$$

But the left hand side of the last inequality is simply $\chi(Z,\theta^m)$ ∎

Lemma 10: If $\theta^m < 1$ and $1 < \sigma \leq \frac{1}{1-\theta^m}$, then in any solution to P(M) we have $a(s^t) > 0$ for all s^t.

Proof. First, consider the case $\sigma > 1$. $a(s^t)$ is greater than zero as long as:

$$\frac{(1-\sigma)}{\chi(Z,\theta^m)}\left[\frac{(\theta^m-1)(1-\alpha)}{1-L(s^t)}+1\right]+\sigma > 0$$

or

$$\frac{(\theta^m-1)(1-\alpha)}{1-L(s^t)}+1 < \frac{\sigma\chi(Z,\theta^m)}{\sigma-1}$$

Since $(\theta^m-1)<0$ the inequality above is indeed true as long as $\sigma \leq \frac{1}{1-\chi(Z,\theta^m)}$. But since by Lemma 9 $\frac{1}{1-\theta^m} < \frac{1}{1-\chi(Z,\theta^m)}$, $\sigma \leq \frac{1}{1-\theta^m}$ is a sufficient condition ∎

Lemma 11(**The Bang-Bang Property:**) Assume $1 < \sigma \leq \frac{1}{1-\theta^m}$. In the solution for the median voter's problem, if there exists $\tilde{t}$ such that the implied tax $\tau_k(s^{\tilde{t}}) < \overline{\tau}$ for all $s^{\tilde{t}}$ then

$$U_C(s^t) = \beta \sum_{s_{t+1}} \Pr(s_{t+1}|s^t) U_C(s^{t+1})[1 + F_k^*(s^{t+1}) - \delta] \;\; \forall \;\; t \geq \tilde{t}$$

and therefore $\tau_k(s^{\tilde{t}}) = 0$ for all $t \geq \tilde{t}$.

We omit the proof here, since it uses the same reasoning as the proof of Lemma 5. The key element of the proof is that the return function in the median voter's problem is increasing in both aggregate consumption and leisure when $1 < \sigma \leq \frac{1}{1-\theta^m}$.

Lemma 12:(**The Capital Tax Result**) Suppose $1 < \sigma \leq \frac{1}{1-\theta^m}$, $\theta^m < E(\theta) = 1$, $k_0(\theta) = \gamma_k + (K_0 - \gamma_k)\theta$ with $K_0 - \gamma_k > 0$, and $b_0(\theta) = \hat{b_0} \; \forall \; \theta \; \in \; \Theta$. Then there exists $\hat{t} > 1$ such that

$$\tau_k(s^t) = \begin{cases} \overline{\tau} = 1 \;\text{ for } t < \hat{t} \\ 0 \leq \tau_k(s^t) < \overline{\tau} \text{ for } t = \hat{t} \\ 0 \text{ for all } s^t \text{ such that } t > \hat{t} \end{cases}$$

The proof is standard, dating from the original work of Chamley (1986). The condition $1 < \sigma \leq \frac{1}{1-\theta^m}$ ensures that the median voter's value function is increasing in aggregate consumption, and therefore it cannot be the case that the constraint UB is always binding when there is discounting. Otherwise the standard reasoning would not apply.

Proposition 12. ***(Labor Tax Result)*** *Suppose $\sigma \leq \frac{1}{1-\theta^m}$ and $\theta^m < E(\theta)$. Then there exists $\hat{t} > 1$ such that, for $t \geq \hat{t}$:*

1. $0 < \tau_l(s^t) < 1$.

2. *$\tau_l(s^t)$ depends on s^t only through $L(s^t)$.*

3. *$\tau_l(s^t)$ is strictly increasing in $[1 - \theta^m]$.*

Case 1: $1 < \sigma \leq \frac{1}{1-\theta^m}$

Because of Lemma 12, and since (NN) is not binding, the first order condition with respect to labor is (for $t \geq \hat{t}$):

$$-\frac{V_{Lt}(Z;\theta)}{V_{ct}(Z;\theta)} = F_L(s^t) \tag{36}$$

In the competitive equilibrium we know that

$$1 - \tau_l(s^t) = -\frac{U_L(s^t)}{F_L(s^t)U_c(s^t)}$$

Combining the last two equations and using (34) and (35) generates

$$1 - \tau_l(s^t) = \frac{a(s^t)}{a(s^t) + b(\theta^m)}$$

Thus, if $1 < \sigma \leq \frac{1}{1-\theta^m}$, by Lemma 10 we have $a(s^t) > 0$, and therefore $0 < 1 - \tau_l(s^t) < 1$ for all s^t.

Case 2: $0 < \sigma < 1$.

Suppose that the constraint UB is not binding for all $t \geq \hat{t}$. Later we will check that constraints. We can write $\tau_l(s^t)$ as:

$$\tau_l(s^t) = \frac{-(\theta^m - 1)}{(\theta^m - 1)[(1-\alpha)(1-\sigma) - 1] + [1 - L(s^t)][(1-\sigma) + \sigma\chi(Z, \theta^m)]} \tag{37}$$

Which implies that $\tau_l(s^t) > 0$ when $0 < \sigma < 1$. In this case, $\tau_l(s^t) < 1$ follows from the intratemporal first order condition in the competitive equilibrium. Otherwise, the marginal productivity of labor should be negative.

Next we claim that, if $\theta^m < 1$ and $0 < \sigma < 1$ then $a(s^t) > 0$ for all s^t.

First, notice that $\tau_l(s^t) < 1$ implies that $a(s^t)$ and $a(s^t) + b(\theta^m)$ must have the same sign. $b(\theta^m) > 0$ with $\tau_l(s^t) > 0$ implies the claim. Finally, since $a(s^t) > 0$ for all $t \geq \hat{t}$, constraint UB is not binding for high enough t ∎

21.7 Numerical Algorithm

Define $U_L(s^t) \equiv \beta^t \Pr(s^t)\left(C(s^t)\alpha(1 - L(s^t))^{1-\alpha}\right)^{1-\sigma} \frac{(1-\alpha)}{1-L(s^t)}$. Also define $V(Z) \equiv \sum_{t,s^t} \beta^t \Pr(s^t) \frac{\left[C(s^t)^\alpha(1-L(s^t))^{1-\alpha}\right]^{1-\sigma}}{1-\sigma}$.

We numerically approximate the solution to the following problem:

$$\max_{\omega_c(\theta^m), C, L, K} \left(\frac{\omega_c(\theta^m)}{\theta^{m\,1-\alpha}}\right)^{1-\sigma} \sum_{t,s^t} \beta^t \Pr(s^t) \frac{\left[C(s^t)^\alpha(1 - L(s^t))^{1-\alpha}\right]^{1-\sigma}}{1-\sigma}$$

$$\text{s.t.} \begin{cases} \omega_c(\theta^m) \leq 1 + \frac{(\theta^m - 1)\sum_{t,s^t} \rho(s^t) U_L(s^t)}{(1-\sigma)V(Z)} \\ \text{Resource constraint} \\ \text{non-negativity constraints} \\ L(s^t) \leq 1 \end{cases}$$

A straightforward extension of Lemma 9 in order to allow for stochastic labor skills shows that, in the solution to the problem above, $\omega_c(\theta^m) \in [\theta^m, 1)$. Let λ be the multiplier related to the first constraint above, which clearly binds in the solution. Let $\xi(s^t)$ be the multiplier on the resource constraint at history s^t. Then the Lagrangean is given by:

$$\mathcal{L} = \sum_{t,s^t} \Pr(s^t)\beta^t \left[\left(\frac{\omega_c(\theta^m)}{{\theta^m}^{1-\alpha}} \right)^{1-\sigma} U(C(s^t), L(s^t)) \right] +$$
$$\lambda \left[(\theta^m - 1) \sum_{t,s^t} \rho(s^t) U_L(s^t) \; + \; (1-\sigma) V(Z)(1 - \omega_c(\theta^m)) \right] +$$
$$\sum_{t,s^t} \xi(s^t) \left[F\Big(L(s^t), K(s^{t-1}), s^t \Big) + (1-\delta)K(s^{t-1}) - C(s^t) - K(s^t) - g(s^t) \right]$$

where $U(C(s^t), L(s^t)) \equiv \frac{\left[C(s^t)^\alpha (1-L(s^t))^{1-\alpha} \right]^{1-\sigma}}{1-\sigma}$.

We then can rewrite $\mathcal{L}$ as:

$$\mathcal{L} = \sum_{t,s^t} \Pr(s^t)\beta^t \left[\left(\frac{\omega_c(\theta^m)}{{\theta^m}^{1-\alpha}} \right)^{1-\sigma} U(C(s^t), L(s^t)) + \right.$$
$$\left. \lambda(1-\sigma) U(C(s^t), L(s^t)) \left(1 - \omega_c(\theta^m) + \rho(s^t)(\theta^m - 1) \frac{1-\alpha}{1 - L(s^t)} \right) \right] +$$
$$\sum_{t,s^t} \xi(s^t) \left[F\Big(L(s^t), K(s^{t-1}), s^t \Big) + (1-\delta)K(s^{t-1}) - C(s^t) - K(s^t) - g(s^t) \right]$$

Taking λ and $\omega_c(\theta^m)$ as given, there exists a functional equation problem (FEP) with a modified return function that solves $\mathcal{L}$ above. Such return function is given by:

$$\widehat{U}(C(s^t), L(s^t); \lambda, \omega_c(\theta^m)) \equiv \left(\frac{\omega_c(\theta^m)}{\theta^{m\,1-\alpha}}\right)^{1-\sigma} U(C(s^t), L(s^t)) +$$
$$\lambda(1-\sigma)U(C(s^t), L(s^t))\left(1 - \omega_c(\theta^m) + \rho(s^t)(\theta^m - 1)\frac{1-\alpha}{1-L(s^t)}\right)$$

Denote $V(K; \lambda, \omega_c(\theta^m))$ the unique function solving the FEP. Using the product topology in the problem in question, we can apply Theorem 3 in Milgrom and Segal (2002). By setting $\frac{\partial V(K;\lambda,\omega)}{\partial \omega_c(\theta^m)} = 0$ we get

$$\omega_c(\theta^m) = [\theta^{(1-\alpha)(1-\sigma)}\lambda]^{-1/\sigma}$$

The numerical solution then uses a two step algorithm. First, for a given λ, and therefore $\omega_c(\theta^m)$ from the equation above, we solve the FEP using value function iteration for a grid of 300 points for the capital stock. In the second step, for each capital stock, we do a grid with 100 points for λ and find $\lambda^*(K)$ that attains $\frac{\partial V(K;\lambda,\omega)}{\partial \lambda} = 0$. Because λ and $\omega_c(\theta^m)$ are related by an equation and $\omega_c(\theta^m) \in [\theta^m, 1)$, we can reduce the size of the grid for $\lambda's$ in a great extent.

We check the numerical solution by evaluating the analytic first-order conditions from the original problem.

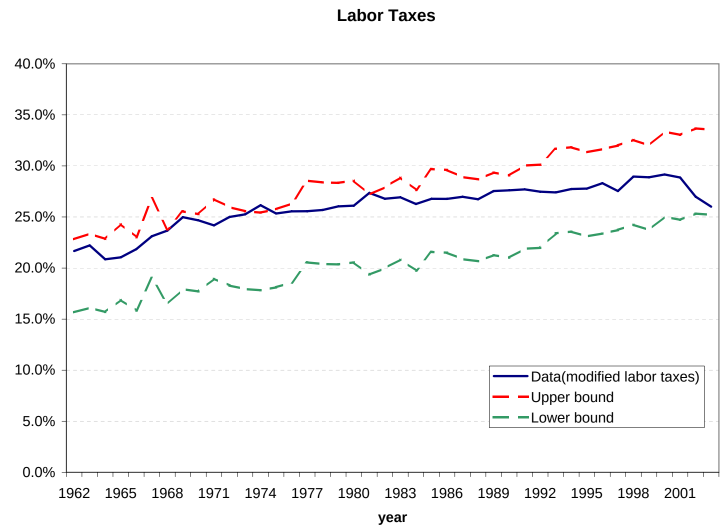

Figure 4: Bounds on labor taxes, $\sigma = 2$.

22 Appendix of Chapter 4

22.1 Proof of Lemma 9

The proof basically works out in two steps. First, we show that we can write beliefs recursively. Second, we use a property about the zeros of analytical functions.

22.1.1 Step 1: Beliefs Updating

Given a pure strategy profile $\sigma_G \in \Sigma_G$ and $h^{g,t} = (R_0, \tau_0, ..., R_{t-1}, \tau_{t-1})$, let $\mu(\cdot)$ be the induced probability distribution over histories for the government. Also let $\mu(|h^{i,t}, y_t^i)$ be the respective conditional probability given history $h^{i,t}$, that is, the probability measure used by agent i in (24) to calculate the expected marginal return of investment. With some abuse of notation, for $h^{i,t}$ consistent with $\hat{h}^{g,t}$, let $f(h^{i,t}|\hat{h}^{g,t}) \equiv \prod_{s=0}^{t-1} f(y_s^i|\hat{R}_s)$.

In addition, for $\hat{h}^{g,t}$ consistent with $\tilde{h}^{i,t}$, with some abuse of notation we can use Bayes' theorem to write:

$$\mu(h^{g,t}|\tilde{h}^{i,t}) = \frac{f(\tilde{h}^{i,t}|\hat{h}^{g,t})\mu(h^{g,t})}{\sum_{\hat{h}^{g,t}} f(\tilde{h}^{i,t}|\hat{h}^{g,t})\mu(\hat{h}^{g,t})} \tag{38}$$

The next lemma shows that $\mu(h^{g,t}|\tilde{h}^{i,t})$ can be written recursively.

Remark: The lemma considers the more general case of mixed strategies $\sigma_{G,t}$: $H^{g,t} \to \Delta(\Upsilon)$.

Lemma 12. *Consider $\sigma_G \in \Sigma_G$. For $\hat{h}^{g,t}$ consistent with $\tilde{h}^{i,t}$ we have that:*

$$\mu(h^{g,t}|\tilde{h}^{i,t}) = P(R_{t-1}|y_{t-1}^i)\frac{\sigma_{G,t-1}(\tau_{t-1}|h^{g,t-1})\mu(h^{g,t-1}|\tilde{h}^{i,t-1})}{\sum_{\tau_{t-1},\hat{h}^{g,t-1}} \sigma_{G,t-1}(\tau_{t-1}|\hat{h}^{g,t-1})\mu(\hat{h}^{g,t-1}|\tilde{h}^{i,t-1})} \tag{39}$$

Proof: First, we have the following:

$$\mu(h^{g,t}) = P(R_{t-1})\sigma_{G,t-1}(\tau_{t-1}|h^{g,t-1})\mu(h^{g,t-1})$$

and

$$f(\tilde{h}^{i,t}|h^{g,t}) = f(y_{t-1}^i|R_{t-1})f(\tilde{h}^{i,t-1}|h^{g,t-1})$$

Using the last two expressions in (38) we get:

$$\mu(h^{g,t}|\tilde{h}^{i,t}) = \frac{f(y_{t-1}^i|R_{t-1})f(\tilde{h}^{i,t-1}|h^{g,t-1})P(R_{t-1})\sigma_{G,t-1}(\tau_{t-1}|h^{g,t-1})\mu(h^{g,t-1})}{\sum_{\hat{h}^{g,t}} f(y_{t-1}^i|R_{t-1})f(\tilde{h}^{i,t-1}|\hat{h}^{g,t-1})P(R_{t-1})\sigma_{G,t-1}(\tau_{t-1}|\hat{h}^{g,t-1})\mu(\hat{h}^{g,t-1})}$$

which yields

$$\mu(h^{g,t}|\tilde{h}^{i,t}) = \frac{f(y^i_{t-1}|R_{t-1})P(R_{t-1})\sigma_{G,t-1}(\tau_{t-1}|h^{g,t-1})f(\tilde{h}^{i,t-1}|h^{g,t-1})\mu(h^{g,t-1})}{\sum_{\hat{R}_{t-1}} f(y^i_{t-1}|\hat{R}_t)P(\hat{R}_{t-1})\sum_{\tau_{t-1},\hat{h}^{g,t-1}}\sigma_{G,t-1}(\tau_{t-1}|\hat{h}^{g,t-1})f(\tilde{h}^{i,t-1}|\hat{h}^{g,t-1})\mu(\hat{h}^{g,t-1})}$$

and

$$\mu(h^{g,t}|\tilde{h}^{i,t}) = P(R_{t-1}|y^i_{t-1})\frac{\sigma_{G,t-1}(\tau_{t-1}|h^{g,t-1})\mu(h^{g,t-1}|\tilde{h}^{i,t-1})}{\sum_{\tau_{t-1},\hat{h}^{g,t-1}}\sigma_{G,t-1}(\tau_{t-1}|\hat{h}^{g,t-1})\mu(\hat{h}^{g,t-1}|\tilde{h}^{i,t-1})} \tag{40}$$

∎

22.1.2 Step 2: Analytical Functions

Lemma 13. *Suppose that $K(y,R)$ is analytic in y for all $R \in \Upsilon$.*

Define

$$m(y) = \int_{\Upsilon} K(y,R)dP(R)$$

and let

$$C = \{y \in [1,\bar{y}] : m(y) = 0\}$$

Then, C has Lebesgue measure zero on $[1,\bar{y}]$.

Proof: We start with a result about analytic functions. Since analytic functions have power-series expansions about all points in their domain, the set of roots is at most countable. The proof is an adaptation of Theorem 10.18 in Rudin (1987) about holomorphic functions.

Claim 1: Suppose $f : [1, \overline{y}] \to \mathbb{R}$ is analytic. Let $Z(f) = \{x \in \mathrm{int}([1, \overline{y}]) : f(x) = 0\}$. Then either $Z(f) = (1, \overline{y})$ or $Z(f)$ has no limit points in $[1, \overline{y}]$. In the latter case $Z(f)$ is at most countable.

Proof of Claim 1: Let M be the set of limit points of $Z(f)$. Take $x_0 \in Z(f)$. We will argue that either $x_0 \in \mathrm{int(M)}$ or x_0 is an isolated point of $Z(f)$. To see this, notice that:

$$f(x) = \sum_{n=0}^{\infty} a_n (x - x_0)^n$$

for $x \in B_r(x_0) \subseteq [1, \overline{y}]$, where $B_r(x_0)$ is an open ball or radius r around x_0.

Then it follows that either all $a_n = 0$, in which case $B_r(x_0) \subseteq M$ and therefore $x_0 \in \mathrm{int}(M)$, or there exists $\hat{n} > 0$ (since $f(x_0) = 0$) such that $a_{\hat{n}} \neq 0$. In the latter case, define:

$$g(x) = \begin{cases} (x - x_0)^{-\hat{n}} f(x) & \text{for } x \in [1, \overline{y}] \setminus \{x_0\} \\ a_{\hat{n}} & \text{for } x = x_0 \end{cases}$$

Because $g(x_0) \neq 0$ and $g(\cdot)$ is continuous, there exists a neighborhood $B_{\tilde{r}}(x_0)$ of x_0 in which $g(\cdot)$ has no zero, and therefore $f(\cdot)$ has no zero in $B_{\tilde{r}}(x_0)$. Then it follows that x_0 is an isolated point of $Z(f)$. This finish the claim that either $x_0 \in \mathrm{int(M)}$ or x_0 is an isolated point of $Z(f)$.

Next take $x \in M$. Because $f(\cdot)$ is continuous, it follows that $x \in M \subseteq Z(f)$. Then either $x \in \mathrm{int(M)}$ or x is a limit point of M. By the reasoning above, x cannot be a limit point of M, because $x \in Z(f)$ and therefore $x \in \text{int(M)}$ or x is an isolated point of $Z(f)$. It then follows that M is open. If $B = [1, \overline{y}] - M$, then B is open since M is the set of limit points of $Z(f)$. Since $[1, \overline{y}]$ is connected, it cannot be the union of the disjoint open sets M and B. Then either $M = (1, \overline{y})$, in which case $Z(f) = (1, \overline{y})$, or $M = \emptyset$. In the latter case $Z(f)$ has at most finitely many points in each compact subset of $[1, \overline{y}]$. But since $[1, \overline{y}]$ is a countable union of compact sets, $Z(f)$ is at most countable ∎

In order to finish the proof, recall that $m(y) = \int_\Upsilon K(y, R) dP(R)$. Since $K(y, R)$ is analytic, let $K(y, R) = \sum_{n=0}^{\infty} c_n(R)(y - y_0)^n$ for some y_0. Notice that the constants $\{c_n(R)\}_n$ depend on R. Then we have:

$$
\begin{aligned}
m(y) &= \int_\Upsilon [K(y, R)]\, dP(R) \\
&= \int_\Upsilon \left[\sum_{n=0}^{\infty} c_n(R)(y - y_0)^n\right] dP(R) \\
&= \sum_{R \in \Upsilon} \left[\sum_{n=0}^{\infty} c_n(R)(y - y_0)^n\right] P(R) \\
&= \sum_{n=0}^{\infty} \left[\int_\Upsilon c_n(R) dP(R)\right] (y - y_0)^n
\end{aligned}
$$

which it is analytic. Then an application of claim 1 implies that set $C = \{y \in [1, \bar{R}] : m(y) = 0\}$ is at most countable, and therefore has Lebesgue measure zero ∎

Given that the composition of analytic functions is itself analytic, a straightforward application of Lemma 13 using (39) in (24) implies that there is a measure zero of agents indifferent between investing or not.

22.2 Proof of Lemma 10

Before we present the proof, consider the following notation. Let $\tau(\hat{R}) \equiv \sigma_G^*(\hat{R})$ and $\tau \equiv \{\tau(R)\}\,|_{R\,\in\Upsilon}$. Given the vector τ, from the individual agent's decision, consider the following function:

$$H(y^i, \tau) = y^i(1 - [E(\tau(R)|y^i)]) - 1 \tag{41}$$

Given assumption 4, the set of agents $i \in I$ such that $H(y^i, \tau) = 0$ is at most countable (see the proof of Lemma 9). Moreover, the set of points y^i such that $H(y^i, \tau) = 0$ is indeed finite as long as there is a state $R \in \Upsilon$ with tax bounded away from the unity. To see this, notice that, for $\hat{y}^i$ high enough, we have $H(\hat{y}^i, \tau) > 0$ whenever $E(\tau(R)|\hat{y}^i) < 1$.

Given the previous reasoning, suppose there are N cutoff points $\{y_i^*\}_{i=1}^N$ satisfying $H(y_i^*, \tau) = 0$. We order them in an ascending order, i.e., $y_i^* \leq y_{i+1}^*$, and let $y_{N+1}^* = \bar{y}$. It is important to keep in mind that since $H(1, \tau) \leq 0$, $\frac{\partial H(y_i^*, \tau)}{\partial y_i^*} > 0$ when i is odd and $\frac{\partial H(y_i^*, \tau)}{\partial y_i^*} < 0$ when i is even.

Notice that, using the implicit function theorem, we have:

$$\begin{aligned}
\frac{\partial y_i^*}{\partial \tau(R)} &= \left[\left(1 - E(\tau(R)|y_i^*) - y_i^* \frac{\partial E(\tau(R)|y_i^*)}{\partial y_i^*}\right)(1 - E(\tau(R)|y_i^*))\right]^{-1} P(R|y_i^*) \\
&= \left(\frac{\partial H}{\partial y_i^*}(1 - E(\tau(R)|y_i^*))\right)^{-1} P(R|y_i^*) \\
&= J(y_i^*)P(R|y_i^*) \qquad (42)
\end{aligned}$$

where $J(y_i^*) \equiv \left(\frac{\partial H}{\partial y_i^*}(1 - E(\tau(R)|y_i^*))\right)^{-1}$.

By the definition of $\{y_i^*\}_{i=1}^N$, the aggregate investment is given by $X(\tau(R), R) = \sum_{i=1}^N \int_{y_i^*}^{y_{i+1}^*} f(y|R)dy$. In a similar fashion aggregate output is $Y(\tau(R), R) = \sum_{i=1}^N \int_{y_i^*}^{y_{i+1}^*} yf(y|R)dy$.

Notice that $\frac{\partial X(\tau(R),R)}{\partial \tau} = \sum_{i=1}^N (-1)^i f(y_i^*|R) \frac{\partial y_i^*}{\partial \tau} < 0$ because $\frac{\partial y_i^*}{\partial \tau} > 0$ when i is odd and $\frac{\partial y_i^*}{\partial \tau} < 0$ when i is even.

In the same way $\frac{\partial Y(\tau(R),R)}{\partial \tau} = \sum_{i=1}^N (-1)^i y_i^* f(y_i^*|R) \frac{\partial y_i^*}{\partial \tau} < 0$ and $\frac{\partial Y(.)}{\partial \tau} - \frac{\partial X(.)}{\partial \tau} < 0$ because $y_i^* \geq 1$ for all i.

We can write the static payoff for the government as:

$$\sum_{R \in \Upsilon} P(R)W(R, \tau(R), \tau(R), Y(\tau(R), R))$$

where

$$W(R, X(\tau(R), R), \tau(R), Y(\tau(R), R)) \equiv (1-\tau(R))Y(\tau(R), R) + (\omega - \tau(R)) + b\tau(R)\ Y(\tau(R), R)$$

Towards a contradiction, suppose that the solution has $\tau_{\bar{R}} < 1$ and $\tau_R > 0$ for some $R \in \Upsilon$. Then consider the following perturbation: increase $\tau_{\bar{R}}$ by $d\tau_{\bar{R}} > 0$ and decreases τ_R by $d\tau_R < 0$ such that it keeps y_N^* fixed.

Then, at the solution, the change Δ in payoff should be zero:

$$\begin{aligned}
\Delta &= (b-1)[P(\bar{R})Y(\bar{R})d\tau_{\bar{R}} + P(R)Y(R)d\tau_R] + \\
&\quad \sum_{i=1}^{N}(-1)^i \left[\sum_{\hat{R}\in\Upsilon} P(\hat{R}) \left([(1+\tau(\hat{R})(b-1))y_i^* - 1] f(y_i^*|R) \right) \right] \left(\frac{\partial y_i^*}{\partial \tau_R} d\tau_R + \frac{\partial y_i^*}{\partial \tau_{\bar{R}}} d\tau_{\bar{R}} \right) \\
&= \Delta_1 + \Delta_2
\end{aligned}$$

where

$$\Delta_1 = (b-1)[P(\bar{R})Y(\bar{R})d\tau_{\bar{R}} + P(R)Y(R)d\tau_R]$$

and

$$\Delta_2 = \sum_{i=1}^{N}(-1)^i \left[\sum_{\hat{R}\in\Upsilon} P(\hat{R}) \left([(1+\tau(\hat{R})(b-1))y_i^* - 1] f(y_i^*|R) \right) \right] \left(\frac{\partial y_i^*}{\partial \tau_R} d\tau_R + \frac{\partial y_i^*}{\partial \tau_{\bar{R}}} d\tau_{\bar{R}} \right)$$

Let the perturbation described above satisfies:

$$\frac{\partial y_N^*}{\partial \tau_R} d\tau_R + \frac{\partial y_N^*}{\partial \tau_{\bar{R}}} d\tau_{\bar{R}} = 0$$

Using (42) we get that:

$$d\tau_R = -\frac{P(\bar{R})f(y_N^*|\bar{R})}{P(R)(f(y_N^*|R)}d\tau_{\bar{R}}$$

For each other y_i^* we have:

$$\frac{\partial y_i^*}{\partial \tau_R}d\tau_R + \frac{\partial y_i^*}{\partial \tau_{\bar{R}}}d\tau_{\bar{R}} = \frac{J(y_i^*)P(\bar{R})}{\sum_{\hat{R}} P(\hat{R})f(y_i^*|\hat{R})}\left[-\frac{f(y_N^*|\bar{R})}{f(y_N^*|R)}f(y_i^*|R) + f(y_i^*|\bar{R})\right]d\tau_{\bar{R}}$$

Assumption 3 implies that $-\frac{f(y_N^*|\bar{R})}{f(y_N^*|R)}f(y_i^*|R) + f(y_i^*|\bar{R}) < 0$ since $y_N^* \geq y_i^*$ for all i. Thus, because $d\tau_{\bar{R}} > 0$, $\frac{\partial y_i^*}{\partial \tau_R}d\tau_R + \frac{\partial y_i^*}{\partial \tau_{\bar{R}}}d\tau_{\bar{R}}$ is negative when i is odd and positive when i is even, and therefore $(-1)^i\left(\frac{\partial y_i^*}{\partial \tau_R}d\tau_R + \frac{\partial y_i^*}{\partial \tau_{\bar{R}}}d\tau_{\bar{R}}\right) \geq 0$ for all i.

Therefore $\Delta_2 > 0$. It remains to show that $\Delta_1 > 0$.

$$\begin{aligned}
\Delta_1 &= (b-1)[P(\bar{R})Y(\bar{R})d\tau_{\bar{R}} + P(R)Y(R)d\tau_R] \\
&= (b-1)d\tau_{\bar{R}}\left[P(\bar{R})Y(\bar{R}) - P(R)Y_L\frac{P(\bar{R})f(y_N^*|\bar{R})}{P(R)f(y_N^*|R)}\right] \\
&= P(\bar{R})(b-1)d\tau_{\bar{R}}\left[Y(\bar{R}) - Y(R)\frac{f(y_N^*|\bar{R})}{f(y_N^*|R)}\right]
\end{aligned}$$

Because $d\tau_{\bar{R}} > 0$, it is sufficient to show that $\frac{Y(\bar{R})}{f(y_N^*|\bar{R})} > \frac{Y(R)}{f(y_N^*|R)}$.Notice that $\frac{Y(\bar{R})}{f(y_N^*|\bar{R})} = \omega\sum_{i=1}^N \int_{y_i^*}^{y_{i+1}^*} y\frac{f(y|\bar{R})}{f(y_N^*|\bar{R})}dy$. and $\frac{Y(R)}{f(y_N^*|R)} = \omega\sum_{i=1}^N \int_{y_i^*}^{y_{i+1}^*} y\frac{f(y|R)}{f(y_N^*|R)}dy$. Each of this variables represents an integral using a normalized probability distribution function $h(y|R') = \frac{f(y|R')}{f(y^*|R')}$ with $h(y^*|R') = 1\ \forall R' \in \Upsilon$.

Hence $\frac{Y(\bar{R})}{f(y_N^*|\bar{R})} > \frac{Y(R)}{f(y_N^*|R)}$ and $\Delta > 0$, a contradiction ∎

22.3 Proof of Proposition 6

Towards a contradiction, fix $\sigma \in \Xi$ with $\sigma_G \in \Xi_G$ being non-trivially essentially private. Define $S = \left\{ \tau \in [0,1] : \tau = \sigma_{G,t}(h^{g,t}, R_t), \text{ for some } R^t \in \Upsilon^t,\ R^{t-1} \subseteq h^{g,t} \right\}$. The fact that σ_G is non-trivially essentially private makes sure that S is not a singleton. Also define

$$H^{*,i} = \{h^{i,t} \in \cup_{t=0}^{\infty} H^{i,t} : \tau_s \in S, \forall s \leq t-1\}$$

Because of the full support condition in assumption 1, all histories in $H^{*,i}$ have non-zero measure.

Since we are considering pure strategies only, using (39), for all $h^{i,t} \in H^{*,i}$ the belief system is given by:

$$\mu(h^{g,t}, R_t|\tilde{h}^{i,t}, y_t^i) = P(R_t|y_t^i) \frac{\mu(h^{g,t-1}|\tilde{h}^{i,t-1})}{\sum_{\hat{h}^{g,t-1}} \mu(\hat{h}^{g,t-1}|\tilde{h}^{i,t-1})}$$

By recursive calculations the above expression can be reduced to:

$$\mu(h^{g,t}, R_t|\tilde{h}^{i,t}, y_t^i) = f(y_t^i|R_t) \prod_{s=0}^{t-1} \frac{f(y_s^i|R_s)}{\sum_{\hat{R}_s} f(y_s^i|\hat{R}_s)} \tag{43}$$

In this way, when σ_G is non-trivially essentially private, beliefs do not depend on the strategy followed by the government. Thus, given the features of σ_G, $\mu(\sigma_G(h^{g,t}, R_t)|h^{i,t}, y_t^i) = \mu(\sigma_G(h^{g,t}, R_t)|\tilde{h}^{i,t}, y_t^i)$ for all $h^{i,t}, \tilde{h}^{i,t} \in H^{*,i}$ and $h^{g,t}$ consistent with histories in $H^{*,i}$.

Now, consider the following one shot deviation strategy $\tilde{\sigma}_G \in \Sigma_G$. Take any period $t > 0$ with history $h^{g,t}$ such that $\sigma_G(h^{g,t}, R_t) < \bar{\tau}^S \equiv \max_\tau\{S\}$ and consider the following alternative strategy.

$$\tilde{\sigma}_G(h^{g,s}, R_s) = \begin{cases} \sigma_G(h^{g,s}, R_s) & \text{if } s \neq t \\ \bar{\tau}^S & \text{if } s = t \end{cases}$$

Let X_t and Y_t be, respectively, the aggregate output and investment generated by $\left\{\sigma^i|_{h^{i,t}}\right\}_{i\in I}$ in time t after R_t is realized. Then $(\sigma|_{h^{g,t}}, \left\{\sigma^i|_{h^{i,t}}\right\}) \in \Sigma$ generates the following payoff:

$$V(\sigma_G|_{h^{g,t}}, \{\sigma_i|_{h^{i,t}}\}_{i\in I}) = (1-\beta)W(R_t, X_t, \sigma_{G,t}(h^{g,t}, R_t), Y_t) +$$
$$\beta V(\sigma_G|_{\{h^{g,t}, R_t, \sigma_G(h^{g,t}, R_t)\ \}}, \sigma_i|_{\{h^{i,t}, \sigma_G(h^{g,t}, R_t)\}})$$

The alternative strategy yields:

$$V(\tilde{\sigma}_G|_{h^{g,t}}, \{\sigma_i|_{h^{i,t}}\}_{i\in I}) = (1-\beta)W(R_t, X_t, \bar{\tau}^S, Y_t) + \beta V(\tilde{\sigma}|_{\{h^{g,t}, R_t, \bar{\tau}^S\}}, \hat{\sigma}^i|_{\{h^{i,t}, \bar{\tau}^S\}})$$

Since $\bar{\tau}^S \in S$, it follows that $\{h^{i,t}, \bar{\tau}^S, y^i_t\} \in H^{*,i}$ and therefore $\hat{\sigma}^i|_{\{h^{i,t}, \sigma_G(h^{g,s}, \tau_s, R_s)\}} = \hat{\sigma}^i|_{\{h^{i,t}, \bar{\tau}^S\}}$ for all $i \in I$.

In addition, from the one shot deviation construction, $\hat{\sigma}|_{\{h^{g,t}, R_t, \sigma_G(h^{g,s}, R_s)\}} = \tilde{\sigma}|_{\{h^{g,t}, R_t, \bar{\tau}^S\}}$.

Hence, $V(\sigma_G|_{\{h^{g,t}, R_t, \sigma_G(h^{g,s}, R_s)\ \}}, \sigma_i|_{\{h^{i,t}, \sigma_G(h^{g,s}, R_s)\}}) = V(\tilde{\sigma}|_{\{h^{g,t}, R_t, \bar{\tau}^S\}}, \hat{\sigma}^i|_{\{h^{i,t}, \bar{\tau}^S\}})$. But since $\bar{\tau}^S > \sigma_G(h^{g,t}, R_t)$, from assumption 2 we have that $V(\tilde{\sigma}_G|_{h^{g,t}}, \{\sigma_i|_{h^{i,t}}\}_{i\in I}) > V(\sigma_G|_{h^{g,t}}, \{\sigma_i|_{h^{i,t}}\}_{i\in I})$ ∎

22.4 Proof of Proposition 7

First, take an arbitrarily agent and history $h^{i,t}$. Suppose that every agent is playing according to σ^{worst}. In this situation the government cannot increase the provision of the public good regardless the tax it chooses. Therefore it is weakly optimal for the government to fully tax investment. In this way, condition (i) in definition 6 is met for all histories. Next, suppose the government is playing according to σ_G^{worst}. Then regardless the individual signals, all the agents assign probability one to full taxation and the optimal action is to choose zero investment. This follows straight from (24). Therefore condition (ii) in definition 6 holds and the proposed strategy it is an equilibrium.

The fact that σ^{worst} yields the the worst equilibrium follows from the observation that the level of provision of the public good is at its minimum value under σ^{worst}. Therefore assumption (2) yields the result ∎

22.5 Proof of Proposition 8

Fix $\mathcal{A}$ an aggregate outcome of $\sigma \in \mathring{\Sigma}$. Condition (1) comes from the definition of aggregating investment decisions over individuals, where such decisions are given by (28) along the stochastic outcome path. Condition (2) comes from both proposition 7 and the definition of the best deviation $W^d(R_t, X(R^t), Y(R^t))$, while (3) follows from condition (ii) in the definition of a perfect Bayesian equilibrium along the equilibrium path.

Conversely, take $\mathcal{A} \equiv \{\tau_t, X_t, Y_t\}_{t=0}^{\infty}$ satisfying the conditions above. Then we construct $\sigma \in \mathring{\Sigma}$ such that $\mathcal{A}$ is the induced aggregate outcome. We construct σ_G as follows. Along the equilibrium path, set $\sigma_{G,t}(\tau^{t-1}(R^{t-1}), R_t) = \tau_t(R^t)$. For all other government's history $\hat{h}^{g,t} \in H^{G,t}$ off-path, set $\sigma_{G,t}(\hat{h}^{g,t}) = \sigma_{G,t}^{\text{worst}}(\hat{h}^{g,t})$. Regarding the individual decisions, for taxes on the path, set decisions as (28), and for all other off-path histories $\hat{h}^{i,t} \in H^{i,t}$, set $\sigma_{i,t}(\hat{h}^{i,t})$=$\sigma_{i,t}^{\text{worst}}(\hat{h}^{i,t})$. Beliefs on path are set according to $E(\tau_t|y^*) = \sum_{R\in\Upsilon}\left(\frac{P(R)f(y^*|R)\tau_t(R_t)}{\sum_{R\in\Upsilon}P(R)f(y^*|R)}\right)$, while beliefs off-path assign full measure to taxes being equal to one.

Next, we check that this strategy profile is indeed an equilibrium. First, by condition (1), households are optimizing giving on the path of play. By construction, on off-path histories optimality also holds. Regarding government optimality, condition (3) prevents the government to obtain a profitable deviation along the path of play when it considers a switch to $\hat{\tau} \in \tau_t(\tau^{t-1}(R^{t-1}), \Upsilon)$. For all other taxes $\hat{\tau} \notin \tau_t(\tau^{t-1}(R^{t-1}), \Upsilon)$, switches from the path are prevented by condition (2). All the equilibrium conditions hold off-path because σ^{worst} is also an equilibrium. ∎

22.6 Proof of Proposition 11

Notice that the best payoff that can be achieved with deterministic tax functions is

$$V^{BP} \leq \sum_{t=0}^{\infty} \beta^t \max_{\tau_t} E[W(\tau_t, R_t| \left\{\sigma_{i,t}^{BP}\right\}_{i\in I})]$$

Define,

$$\tau^* = argmax_{\tau_t} E[W(\tau_t, R_t| \left\{\sigma_{i,t}^{BP}\right\}_{i\in I})]$$

And notice that

$$E[W(\tau^*, R_t| \left\{\sigma_{i,t}^{BP}\right\}_{i\in I})] < E[W(\sigma_G^*(R_t), R_t| \left\{\sigma_{i,t}^{*}\right\}_{i\in I})] \tag{44}$$

otherwise it would contradict the Ramsey solution.

Then, from (44), we obtain:

$$V^{BP} \leq \sum_{t=0}^{\infty} \beta^t E[W(\tau^*, R_t|x_{\forall i\in I}^{*i})] < \sum_{t=0}^{\infty} \beta^t E[W(\sigma^*(R_t), R_t| \left\{\sigma_{i,t}^{*}\right\}_{i\in I})] = V(\sigma^*)$$

Since (44) holds regardless the magnitude of β, the result holds for $\beta \in [\hat{\beta}, 1)$, the sufficient condition under which the best equilibrium can be achieved using public strategies ∎