EINAUDI INSTITUTE FOR ECONOMICS AND FINANCE

EIEF Working Paper 02/10

January 2010

Heterogeneous Labor Skills, The Median Voter and Labor Taxes

by

Facundo Piguillem

(EIEF)

Anderson L. Schneider

(University of Minnesota and FRB Minneapolis)

Heterogeneous Labor Skills, The Median Voter and Labor Taxes*

Facundo Piguillem

University of Minnesota
and FRB Minneapolis

Anderson L. Schneider †

University of Minnesota
and FRB Minneapolis

November 17, 2009

Abstract

We explore 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, we model this relation through a political economy channel by developing a median voter result over sequences of capital and labor taxes. We study an infinite horizon economy in which agents are heterogeneous with respect to both initial wealth and labor skills. The first result is the following: if initial capital holdings are an affine function of skills, then the median agent is decisive in any pairwise majority rule election. 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. Finally, numerical simulations show that temporary increases in inequality could imply either higher or lower labor taxes, depending on the sign of the correlation between inequality and TFP. This finding provides a simple an intuitive answer to why fiscal policy is procyclical in some countries. When calibrated for the US economy, the model does a good job on fitting both the increasing trend and the levels of labor taxes in the last decades, and on matching short run co-movements.

Keywords: Median Voter. Business Cycle. Labor Taxes. Procyclical Fiscal Policy.

---

*The authors are grateful to Larry Jones and Narayana Kocherlakota for their support and encouragement. We would also like to thank V. V. Chari, Christopher Phelan, Víctor Ríos Rull, Pierre Yared, seminar participants at Minneapolis FRB and workshop participants at the University of Minnesota for comments. Remaining mistakes are ours. The views expressed herein are those of the authors and not necessarily those of the Federal Reserve Bank of Minneapolis or the Federal Reserve System.

†Contact Schneider: Department of Economics, University of Minnesota, 4-101 Hanson Hall, 1925 Fourth Street South, Minneapolis, MN 55455. Email: `als@econ.umn.edu`

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. It is reasonable to think that these properties would be potentially affected by the assumptions on the politico-economic process. In this sense, it is desirable to construct a model in which the impact of these assumptions is minimized. A natural way to achieve this would be 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 Majority Rule. 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 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. 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. A natural question then arises: is there an equilibrium policy? 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. Further, we can show, although is not present in this paper, that these results can be generalize when sequential voting (lack of commitment) is allowed, using the same equilibrium definition as in Bernheim and Slavov (2008).

The proof of the existence of the Condorcet winner 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, in the non-stochastic steady state marginal labor taxes depend directly on the skewness of the distribution. In this case, the most interesting feature is that along the business cycle the correlation between labor taxes and output is ambiguous. We show a numerical example that can generate either procyclical or counter-cyclical taxes, depending on how the dispersion of the distribution of individual productivities changes along the business cycle. The intuition is the following: along the business cycle the median voter balances her desire for "inefficiency smoothing" and "redistribution smoothing". If the correlation between inequality and efficiency (TFP) is positive both incentives are aligned and labor taxes are counter-cyclical as in the traditional Ramsey approach. However, if the correlation between inequality and efficiency is negative and large enough in absolute value, the redistribution effect dominates and marginal labor taxes are procyclical.

We further show in an extend economy with endogenous government spending (public good) how this mechanism provides an answer to the question posed among others by Alesina et al. (2008) and Ilzetzki and Vegh (2008), related to the empirical observation that fiscal policy is often procyclical in developing countries and counter-cyclical in developed ones. If the public good (government consumption) is a normal good, it should by definition be procyclical, consuming (spending) more in good times and less in bad times. This coupled with a procyclical tax policy would immediately generate the result for almost all countries. However, the measures generally used empirically are normalized with respect to the GDP. In this case, the correlation between government spending (over GDP) and GDP is very small in absolute value, making the tax policy the dominant component. As a result, economies with small enough changes in inequality along the business cycle will exhibit counter-cyclical fiscal policies, while economies with large increases of inequality in recensions (and large decreases in booms) will exhibit procyclical fiscal policies.

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. In this context, the example about the procyclicality of fiscal policy should make this point very clear. There is no need to assume any kind of specific imperfection in the institutions to get the result, depending on the underlaying stochastic process, procyclical fiscal policy could very well be the desire outcome for rational agents that choose in an environment with full commitment.

As we mentioned before, our model gives an extension to the median voter result presented in B & B.[1] Other than the already highlighted differences in the characteristics of the

[1] 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.

physical environment, our results depend on the assumption of balanced growth preferences defined over consumption and 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 B & B, 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.[2]

[2] 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.

Regarding the empirical results, Chari, Kehoe and Christiano (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 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 of the existence of a Condorcet winner. Section V characterizes the Condorcet winner, while section VI considers stochastic skills. Section VII shows the numerical results and the last section concludes. Appendix 2 contains the generalization of the median voter result when there is no commitment.

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.[3]

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

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

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

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 $\sum_{t=0}^{\infty} \beta^t u\Bigg(c(s^t, \theta), 1 - l(s^t, \theta)/\theta\Bigg)$, 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:

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

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:

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

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.[4] 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 order conditions with respect to individual consumption and output yield:

$$\frac{\alpha\beta^t \Pr(s^t|s_0)}{c(s^t;\theta)} = p(s^t)\lambda(\theta) \tag{5}$$

$$\frac{(1-\alpha)\beta^t \Pr(s^t|s_0)}{\theta - l(s^t;\theta)} = p(s^t)(1-\tau_l(s^t))w_t(s^t)\lambda(\theta) \tag{6}$$

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:

$$p(s^t) = \frac{E(\varphi)\alpha\beta^t \Pr(s^t|s_0)}{C(s^t)} \tag{7}$$

$$p(s^t)(1-\tau_l(s^t))w_t(s^t) = \frac{E(\varphi)(1-\alpha)\beta^t \Pr(s^t|s_0)}{1-L(s^t)} \tag{8}$$

We normalize the initial price p_0 so that $E(\varphi) = 1$.[5] Then individual allocations can be written as:

$$c(s^t;\theta) = \varphi(\theta)C(s^t) \tag{9}$$

$$1 - l(s^t;\theta)/\theta = \varphi(\theta)\theta^{-1}[1-L(s^t)] \tag{10}$$

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

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

[4] 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.

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

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

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

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:

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

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* $\varphi : Z^\infty \times \Theta \rightarrow \mathbb{R}_+$ *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 Existence of Equilibrium: the Condorcet winner

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:

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

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

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

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:

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

Therefore the sign of the derivative does not depend on θ $\square$

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.[6]

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

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

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 $\underline{\theta}$.[7]

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.[8]

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

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

[8] 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:

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

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

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 $\square$

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*

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

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:*

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

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:

$$\left[\frac{(1-\alpha)(\theta^m - E(\theta))\beta^t \Pr(s^t)}{\varphi(\theta^m)[1-L(s^t)]^2}\right] - \frac{(1-\alpha)\beta^t \Pr(s^t)}{[1-L(s^t)]} + \lambda(s^t) F_L(s^t) = 0 \quad \text{for } t \geq 1 \tag{14}$$

The implied tax on labor is given by:

$$1 - \tau_l(s^t) = \left[\left(\frac{(1-\theta^m)}{\varphi(\theta^m)}\right)\frac{1}{1-L(s^t)} + 1\right]^{-1} \tag{15}$$

Let $H = \frac{(1-\theta^m)}{\varphi(\theta^m)} > 0$. Then (15) can be rewritten as:

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

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

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.[9]

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 for all $t > \hat{t}$, for some $\hat{t} \geq 2$. In addition, the statements are weaker in the sense that they depend on

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

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:

$$\tau_l(s^t) = \frac{(1-\theta^m)}{(\theta^m-1)[(1-\alpha)(1-\sigma)-1]+[1-L(s^t)][(1-\sigma)+\sigma\chi(Z,\theta^m)]} \tag{16}$$

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:

$$\varphi(Z;\theta_0^i) = 1 + (1-\beta)\left[\left(\widetilde{W}_0(\theta_0^i, T, \tau_{k0}) - \mathrm{E}(\widetilde{W}_0(\theta_0^i, T, \tau_{k0}))\right) + (\theta_0^i - 1)\sum_{s^t} \rho(s^t) U_L(s^t)\right] \tag{17}$$

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:

$$\tau_l(s^t) = \frac{(1-\theta_0^m)\rho(s^t)}{\varphi(\theta_0^m)(1-L(s^t)) + (1-\theta_0^m)\rho(s^t)} \tag{18}$$

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 (18) 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 $s = low$.

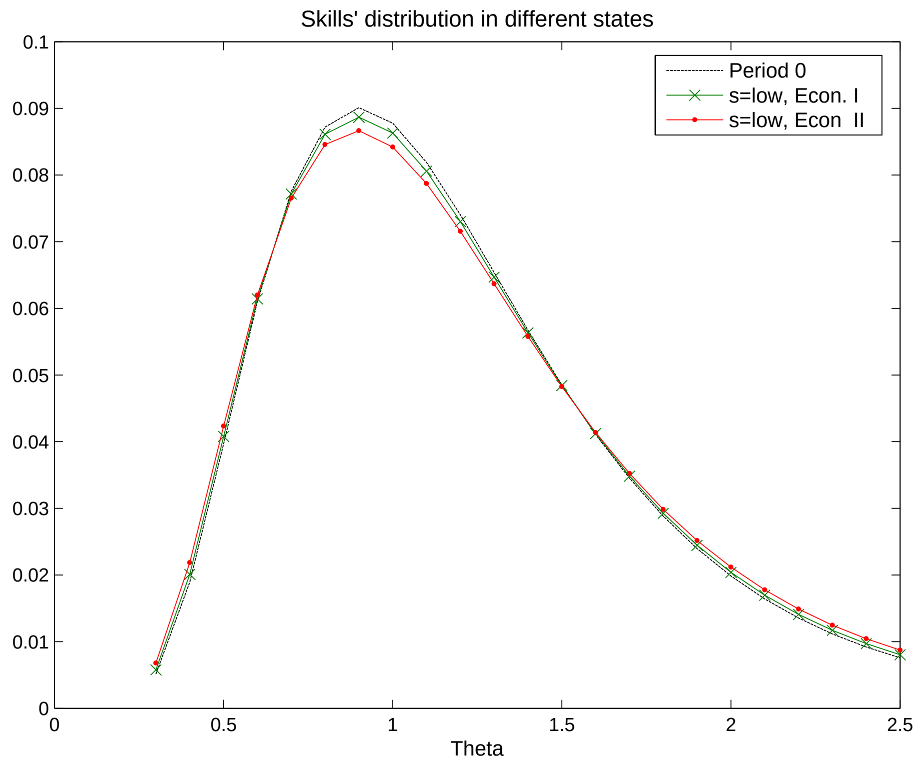

Figure 1: Skill distribution in period zero and in state $s = low$.

Next we show the calculated taxes for each state.

State	Economy I τ	Economy I L	Economy II τ	Economy II 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.

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 (16), 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 US(data from 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).[10] 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=$\{s1 = (A_H, \rho_H), s2 = (A_H, \rho_L), s3 = (A_L, \rho_H), s4 = (A_L, \rho_L)\}$.

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

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 $\alpha = 0.38$.[11]

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

[11] 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.

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.

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.

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

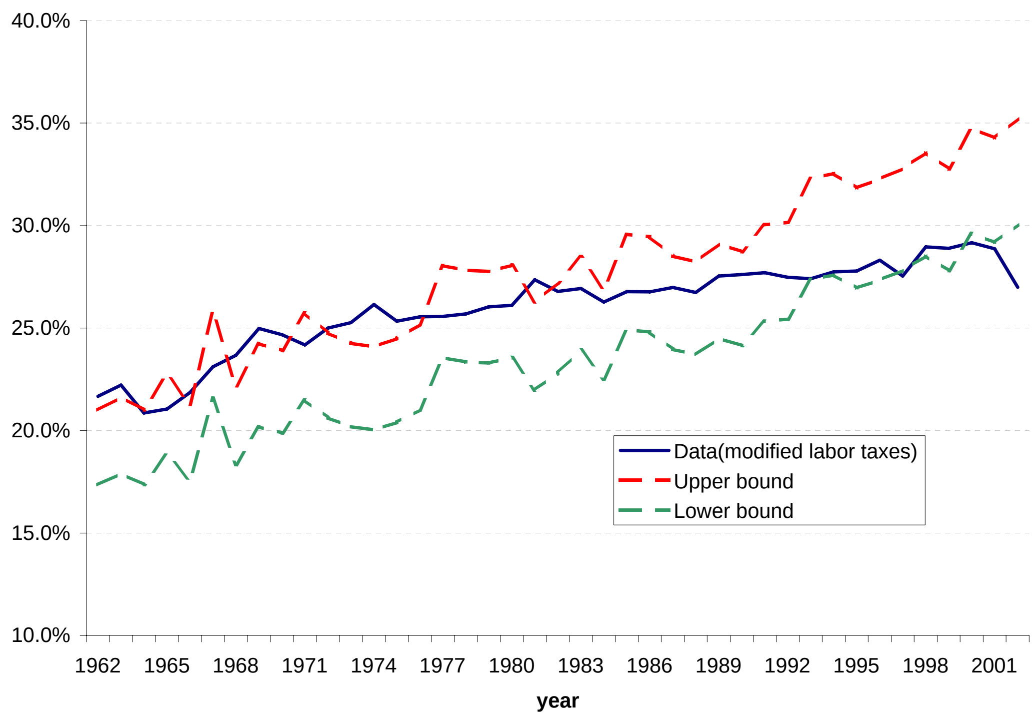

Figure 3: Bounds on average labor taxes in the calibrated model.

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[12], the model accounts for about two times 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, Kehoe and Christiano (1994). When we feed in 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. In the last line we present a sensibility exercise using an utility function with $\sigma = 2$ instead of logarithmic. All the remainder parameters remain the same except by the variance of θ that is set to match the standard deviation of output growth. As we can see in the last line of Table 4, the switch in the sign of the correlations still happens but now in less extreme way. In this case the generated correlations are even closer to those observed in the data, while that the volatilities of both the tax and hours worked are smaller, not only than those in the data but also than the ones generated by the model with logarithmic utility function.

[12]This specific period does not include the significant decrease in average labor taxes between 2002 and 2003.

	Statistics				Correlations		
	St.Dev(ΔY)	St.Dev(L)	Mean Tax	St.Dev(τ)	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	0.25	0.001	0.7	0.2	1
Model changing types (sigma=1)	0.02	0.006	0.25	0.009	-0.6	-0.2	-1
Model Changing types (sigma=2)	0.02	0.007	0.26	0.006	-0.4	-0.08	-0.4

Table 4: Selected statistical properties of the model.

7.2 Example: The optimality of procyclical fiscal policies

As we mentioned before, recently there has been a debate about the procyclicality of fiscal policies in Developing Countries, especially Alesina et al. (2008) and Ilzetzki & Vegh (2008). This is considered a puzzle because is understood that optimal fiscal policies should be counter-cyclical. On the contrary, the numerical simulations of Table 4 imply that this might not necessarily be true. Once the constraint that political decisions have to satisfy the media agent in the economy is included in the definition of optimality the sign of the correlation could reverse. That is, the puzzle would be why fiscal policy is counter-cyclical in developed countries. To shed some light on this issue we solve an economy with endogenous government spending. Specifically we modify the agent's utility function including a public good G, that is, the new utility function is: $U(c(\theta), l(\theta)) + \gamma log(G)$. With this utility function all the theoretical results still hold. This slight modification to the baseline model allow us to compute variables that have a closer empirical counterpart. In the literature the procyclicality or countercylicality of the fiscal policy is defined according to the correlation between the fiscal surplus (or deficit) and some measure of GDP gap. Where (primary) surplus is define as

$$Surplus = Tax\ \ revenues - Government\ \ spending - transfers$$

In our economy we define surplus as

$$Surplus_t = L_t W_t \tau_{lt} - G_t + T_t$$

Thus, on the one hand, if the correlation between $Surplus_t$ (or $Surplus_t/Y_t$) and Y_t is positive fiscal policy would be countercyclical (increase the surplus -save- in good times and decrease the surplus -borrow- in bad times). On the other hand, if the correlation is negative the fiscal policy would procyclical (decrease the surplus -borrow - in good times and increase the surplus -save - in bad times).

The strategy the solve this modified version of the model is the same as before. In addition, we use the same parameters as in the numerical simulations (with logarithmic utility function). It rest to calibrate the additional parameter γ. We set $\gamma = 0.2\alpha$, this makes sure that in average government consumption is 14% of GDP. Moreover, there is an issue about the per period transfers. As is clear from Section 3 the Ricardian Equivalence in this economy holds, therefore what matters is only the total amount of transfers T and not its distribution over time, T_t. We assume that transfers are equally distributed over time, that is $T_t = (1-\beta)T$. Once the model is solved we simulated 1000 times, with the initial level of capital set at the non stochastic steady state. Then, we constructed the variable surplus as defined above and we computed several correlations that are shown in Table 5.

Variable	Correlation with Y
Labor Tax	-0.54
G	0.62
G/Y	-0.03
Surplus	-0.76
Tax Revenues	-0.72
Surplus/Y	-0.73
Tax Revenues/Y	-0.77

Table 5: Procyclical Fiscal Policy.

First, notice that not only the tax rate on labor incomes is procyclical but also the total revenues from labor taxation. Second, since the public good is a normal good the government spending in the public good is procyclical as well. As a result, the surplus as a whole is procyclical. However, empirically this measures are generally computed as a proportion of the GDP, as we can see from Table 5 the procyclicality of fiscal policies is still true in this case. The correlation is slightly smaller (in absolute value) but still highly negative. The main difference is in the correlation of the public good, which now exhibits a correlation very close to zero, is due to the fact that, as consumption, the public good is less volatile than GDP. Thus, the sign of the correlation between fiscal surplus and GDP depends almost entirely on the correlation between revenues (over GDP) and GDP. Therefore, less variance in the inequality ratio or a small enough correlation between the inequality ratio and GDP would generate optimal countercyclical fiscal policies. It is worth to mention that these results still hold when instead of using fiscal surplus as defined above we use the change in government debt holdings by the agents computed as in Chari et al. (1999), which implies that the correlation is robust to the assumption about per period transfers. Finally, if we assume that the economy is not in steady state but converging to it from some initial level of capital stock, as it seems to be the case in a significative sample of economies, the initial level of capital stock does affect the size and potentially the sign of the correlations. Thus, economies that converge from bellow exhibit a stronger negative sign while economies that are converging from above generate a correlations closer to zero and even positive in some cases. We do not present those results in here because we think that those issues goes beyond the scope of this paper and they should address in future research.

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 et al. (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 model also matches the negative sign of the comovements between taxes and labor and output. In addition, we show by means of an example how this approach can be useful explain potential empirical puzzles" like the observed procyclicality of fiscal policies in developing countries. In our model this comes naturally: the demand for more redistribution in bad times together with the fact that the public good is normal generates the result.

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.

References

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

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

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.

Azzimonti, M., E. De Francisco, and P. Krusell (2008): "Aggregation and Aggregation ," *Journal of the European Economic Association*, 6, 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., and S. N. Slavov (2009): "A Solution Concept for Majority Rule in Dynamic Settings," *Review of Economic Studies*, 76, 33–62.

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

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

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

Chari, V., and P. J. Kehoe (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 (2009): "Politico-economic consequences of rising wage inequality ," *Journal of Monetary Economics*, 56, 43–61.

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

Gans, J., and M. Smart (1996): "Majority voting with single-crossing preferences," *Journal of Public Economics*, 59, 219.

Heathcote, J., K. Storesletten, and G. Violante (2008): "The Macroeconomic Implications of Rising Wage Inequality in the United States," *NBER Working Paper*, 14052.

Ilzetzki, E. and, V. C. (2008): "Procyclical Fiscal Policy in Developing Countries: Truth or Fiction?," *NBER Working Paper*.

Inman, R. (1987): "Markets, Government, and the 'New' Political Economy," *Handbook of Public Economics*.

Kehoe, T., and D. Levine (2001): "Liquidity Constrained Markets versus Debt Constrained Markets," *Econometrica*.

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

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

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*.

Milgrom, P., and I. Segal (2002): "Envelope Theorems for Arbitrary Choice Sets," *Econometrica*, 70(2).

Mirrlees, J. (1971): "An Exploration in the Theory of Optimum Income Taxation," *Review of Economic Studies*, 38, 175.

Ohanian, L., A. Raffo, and R. Rogerson (2006): "Long-Term Changes in Labor Supply and Taxes: Evidence from OECD Countries, 1956-2004," *FRB of Kansas City Working Paper No. 06-16*.

Persson, T., and G. Tabellini (1994): "Is Inequality Harmful for Growth?," *American Economic Review*, 84, 600.

Phelan, C., and E. Stacchetti (2001): "Sequential Equilibrium in Ramsey Tax Model ," *Econometrica*, 69, 1491–1518.

Piguillem, F., and A. L. Schneider (2007): "A note on redistribution, optimal fiscal policy and corner solutions," *(www.econ.umn.edu/ ~ als/Research/Taxescornersol.pdf)*.

Werning, I. (2006): "Optimal Fiscal Policy with Redistirbution," *Quarterly Journal of Economics*.

Yared, P. (2007): "Politicians, Taxes, and Debt," *Mimeo, MIT*.

A Appendix 1

A.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:

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

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:

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

or

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

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:

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

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

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 $l \in [0,1]$:

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

The transversality condition (Tvc) $\lim_{t\to\infty}\sum_{s^t} p(s^t)k(s^t;\theta) = 0$ 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 $w(s^t) = F_L(s^t)$ ■

A.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 $\widehat{\theta} \equiv \sup\,\{\underline{\theta}(Z) : Z \in \tilde{\tilde{\Xi}}\}$. 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 $L(s^t) \subseteq Z)$ ■

A.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:

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

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

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

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

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

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

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

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

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 $\square$

A.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 $\{C^*(s^t), K^*(s^t)\}_{s^t \not> s^{\tilde{t}}}$ is a feasible allocation for the median voter's problem that improves the objective function, a contradiction $\square$

A.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:

$$\frac{(1-\theta^m)}{\varphi(\theta^m)} = \mu \frac{C_0}{\alpha(K_{-1} - \gamma_k)} \frac{1}{(F_{k_0} - \delta)} \tag{19}$$

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:

$$\frac{1}{C_0^2} R_0 (K_{-1} - \gamma_k) \frac{(1-\theta^m)}{\varphi(\theta^m)} - E\left[\frac{\bar{\tau}(F_k(s^1) - \delta)}{C(s^1)} \mid s_0\right] \leq 0 \tag{20}$$

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 (19) 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 □

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

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

Let $UL = \sum_t \beta^t\Pr(s^t)\left(C(s^t)^\alpha(1-L(s^t))^{1-\alpha}\right)^{1-\sigma}\frac{(1-\alpha)}{1-L(s^t)}$. 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:

$$\omega_c(\theta) = \frac{[U_{c0}\widetilde{W}_0(\theta, T, \tau_{k0}) + \theta UL]}{(1-\sigma)V(Z)} \tag{21}$$

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:

$$V(Z, \theta^m) = \frac{[\chi(Z, \theta^m)]^{1-\sigma}}{\theta^{(1-\alpha)(1-\sigma)}} V(Z) \tag{22}$$

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

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

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 □

A.6.1 Characterization

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

$$V(Z, \theta^m) = V(Z)\left\{\frac{u_{c0}[\widetilde{W}_0(\theta^m, T, \tau_{k0}) - E[\widetilde{W}_0(\theta, T, \tau_{k0})]] + (\theta^m - 1)UL + (1-\sigma)V(Z)}{[(1-\sigma)V(Z)]}\right\}^{1-\sigma} \tag{23}$$

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:

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

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

We use the notation of equation (22) 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:

$$\hat{V}_C(s^t) = \beta^t \Pr(s^t)[\chi(Z, \theta^m)]^{1-\sigma} \left\{ \frac{(1-\sigma)}{\chi(Z, \theta^m)} \left[\frac{(\theta^m - 1)(1-\alpha)}{1 - L(s^t)} + 1 \right] + \sigma \right\} U_c(s^t) \tag{24}$$

$$\hat{V}_L(s^t) = \beta^t \Pr(s^t)[\chi(Z, \theta^m)]^{1-\sigma} \left\{ \frac{(\theta^m - 1)[(1-\alpha)(1-\sigma) - 1]}{[1 - L(s^t)]\chi(Z, \theta^m)} + \frac{1-\sigma}{\chi(Z, \theta^m)} + \sigma \right\} U_L(s^t) \tag{25}$$

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:

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

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 4. ***(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 $t \geq \hat{t}$):

$$-\frac{V_{Lt}(Z;\theta)}{V_{ct}(Z;\theta)} = F_L(s^t) \tag{26}$$

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 (24) and (25) 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:

$$\tau_l(s^t) = \frac{-(\theta^m-1)}{(\theta^m-1)[(1-\alpha)(1-\sigma)-1]+[1-L(s^t)][(1-\sigma)+\sigma\chi(Z,\theta^m)]} \tag{27}$$

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 □

A.7 Numerical Algorithm

We numerically approximate the solution to the following problem:

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

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

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

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

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:

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

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.