Redistribution with Labor Markets Frictions*

Carlos E. da Costa† Lucas J. Maestri‡ Marcelo R. Santos§

VERY, VERY PRELIMINARY

Abstract

How does the presence of labor market frictions affect optimal redistributive policies? Embedding an otherwise standard Mirrlees' economy in a directed search environment, we characterize constrained efficient allocations and show that distortions in labor supply the intensive and extensive margins always share the same sign. In contrast with the no frictions economy, incentive compatibility does not imply monotonicity in effort. We show that efficiency does imply. We ask whether labor income tax schedules suffice to implement constrained allocations. After showing that it does not, we derive a policy that: i) improves upon the current US system, and; ii) cannot be improved by local reforms. We quantify the gains from moving to such policy. ***Keywords:*** *Mirrlees' problem; Directed Search.* ***JEL Classification:*** *D82, H21.*

LABOR market policies typically combine non-linear income taxes with unemployment benefits to jointly address redistributive and insurance concerns of society. However useful these instruments are for improving the distribution of income across agents and/or across states of nature, they are also characterized by adverse incentive effects. When labor markets are not frictionless the possibility of unemployment arises and unemployment benefits are used to alleviate its consequences. A side effect of such policy is to influence the type of jobs to which workers apply and if they apply at all. By distributing the burden of financing the state across all agents labor income taxes play

*Mestri thanks National Council for Scientific and Technological Development CNPq -, and da Costa thanks CNPq project 301140/2017-0 for financial support. This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Fiance Code 001. All errors are our responsibility.

†EPGE FGV Brazilian School of Economics and Finance

‡EPGE FGV Brazilian School of Economics and Finance

§Insper

an important redistributive role. Yet, they also influence unemployment through their impact on the relative attractiveness of different jobs.

Despite their close connection, these two aspects of labor market policies are seldom jointly studied. Whereas the former is normally assessed under the assumption that labor markets are frictionless, thus following Mirrlees' (1971) mechanism design tradition, the latter advances by abstracting from heterogeneity in workers' productivities and focusing on potential labor market inefficiencies. In this paper, we depart from this tradition and focus on the interaction between redistribution and insurance with frictional labor markets.

We take into account the fact that trade in labor market is decentralized and frictional; as in Golosov et al. (2013) the labor market is characterized by competitive search. In contrast with Golosov et al. (2013) we assume that heterogeneity is on the side of workers while firms are identical. This brings our work closer to Mirrlees (1971) while allowing us to including the consequences for unemployment, which is absent from most discussions of redistributive policies. More importantly, following the mechanism design approach, our focus is on allocations. This allows us to ask, among other things, whether the two policy instruments aforementioned, income taxes and unemployment benefits, make the best use of all publicly available information.

In our model, firms post vacancies for jobs described by an output to be produced by the worker and the earnings they are entitled to if and only if this output is produced. Posting a vacancy is costly for the firm not for the worker as in Golosov et al. (2013). Once a vacancy is posted, the probability that it is filled depends on labor market tightness. Workers apply to vacancies that maximize their expected utility. A matching function summarizes the process which brings together firms and workers interested in the same labor contracts. Not all workers willing to accept a posted contract find a job, and not all vacancies with interested candidates are filled.

If an allocation is efficient: *i)* effort and the probability of finding a job increase with productivity; *ii)* wedges on effort and vacancies share the same sign, and; *iii)* there are distortions at the bottom even when the underlying (Paretian) objective entails finite inequality aversion.[1] Property *i* provides a straightforward efficiency test which does not require the observation of policy instruments.

Next, we ask whether an informationally feasible unemployment benefit along with a non-linear income tax schedule suffice to implement optimal allocations. We find it *not*

[1] Seade (1977) proves that, in a Mirrlees', as long as the planner's objective assigns non-negative but finite weight on all agents, marginal tax rates are zero at the bottom of the distribution unless there is bunching. We find positive wedges at the bottom for a Utilitarian objective.

to be the case and show how tax schedules that use information on the output produced by the agent can implement the optimum.

While earnings are easily observed, at least in developed countries, implementation of constrained efficient allocations requires either job postings or agents' outputs to be observed as well. Within the model, job postings are public information used by workers to decide which jobs to apply to. In practice, such information may not be so easy to gather and process. Along the same lines recovering each worker's output from the revenues generated by firms may not be feasible in multi-worker firms. These practical information costs begs the question of what is lost when we rely on a tax system comprised of unemployment benefits and labor income taxes only.

To answer this question we approximate the current US schedule with the parametric specification of Musgrave (1959); Bénabou (2000, 2002) and obtain an assessment of the utility attained by workers of each productivity level. We then derive a constrained efficient allocation that delivers this utility profile while satisfying all the first order conditions that characterize efficiency. Because the program is not concave this may not be the optimum. Still, when we compare the resources used in each case we find that there are gains from replacing the current policy with the latter. These gains are non-negligible but may not be large enough to justify the costs of gathering the required information for implementing them.

The rest of the paper is organized as follows. Following a brief literature review, we describe in Section 1 the model economy. The program used to characterize efficient allocations is presented in Section 2 where we also present its solution. In Section 3 we show that a non-linear income tax schedule alone cannot implement the constrained efficient allocation, and provide an example of informationally feasible instruments that implements such allocations. Finally, in Section 4 we assess the cost for the US government of not having (or not using) such richer instruments. Section 5 concludes. All proofs are in the appendix.

Literature Review

In Mirrlees' (1971) seminal work, a characterization of constrained efficient allocations is offered in a world where agents with private information about their productivities must be offered incentives for truthfully revealing them. This mechanism design approach to redistribution policies which has been dominant until the end of the 20th century was often criticized for its oversimplified environment which precluded the discussion of important aspects of labor market policies. Among them, the nonexistence of involuntary unemployment.[2] In this paper we try to fill this gap.

Ours is not the first work to consider taxation in the presence of labor market frictions. As early as Pissarides (1985), the interaction between taxes and labor market frictions is studied in an equilibrium unemployment model. More recently Lehmann et al. (2006, 2011); Golosov et al. (2013); Schaal and Taschereau-Dumuouchel (2014); Lehmann et al. (2016); Kroft et al. (2017) are some of the works that have dealt with this issue. These works abstract from intensive margin choices, while introducing novel margins of response and/or heterogeneity. Whether a random search or directed search is considered is also an important difference across works.

Our focus on the consequences of labor market frictions for *redistributive* policies, means that we refrain from addressing the type of ex-post wage dispersion and residual inequality that motivates Golosov et al. (2013). We do so not because we think this aspect is not important, but because it has already been competently addressed in Golosov et al. (2013). Our contribution is, in this sense, complementary to theirs.

In what we view as one of the most innovative contributions to the area, Lehmann et al. (2006), apply a mechanism design approach to an optimal taxation problem in a labor market characterized by random search. The assumption of ex-post wage bargain leads to surpluses for both the firm and the worker. The mechanism must extract information from the pair which, however cooperative, still have conflicting interests. Only extensive margin responses are considered. We, in contrast, make a minimum departure from Mirrlees's classic work to assess how labor *market frictions alone* affect optimal tax prescriptions and create extensive margin distortions. Moreover, we consider competitive search, which eliminates all surplus beyond the informational rents that accrue to workers.

Some of the aforementioned works also rely on a perturbation/sufficient statistic approach, making it hard to assess whether all relevant policy instruments are used. We, in contrast, follow a mechanism design approach which allows us to assess what is feasible under the informational restrictions that real world governments face.

Closest to our work is Boadway and Cuff (2016). They too combine insurance and redistribution concerns in policy design by assuming workers heterogeneity. They do not consider intensive margin adjustments but allow for non-participation by assuming that agents are heterogeneous in their participation costs.[3] Adding another dimension of heterogeneity would raise all the well known issues associated with multi-dimensional

[2] In some works, a fraction of workers does not work, e.g., Jacquet et al. (2013), but non-employment is voluntary.

[3] Non-participation takes also center stage in Kroft et al. (2017).

screening, which is why we refrain from doing so. In the spirit of directed search we also assume that firms can commit to posted wage, while wages are determined through bargain in Boadway and Cuff (2016).

Finally, Geromichalos (2015) studies how taxes used to finance an unemployment insurance program affects efficiency. As in our case, a directed search model is used. Contrary to what we do here, Geromichalos (2015) assumes that workers are identical and focuses on the externalities created by the UI funding itself.

1 Environment

The consumer/worker side of the economy is as in Mirrlees (1971). There is a continuum of agents with preferences defined over consumption, c, and effort, n, represented by

$$U(c,n,\theta) := \varphi(c) - \theta\eta(n),$$

where $\varphi : \mathbb{R}_+ \mapsto \mathbb{R}$ is a smooth, increasing, and strictly concave function and $\eta : \mathbb{R}_+ \mapsto \mathbb{R}$ is a smooth, increasing, and strictly convex function.

Heterogeneity across agents is captured by the disutility of effort parameter θ, distributed in the compact interval $[\underline{\theta}, \bar{\theta}]$ with $\underline{\theta} > 0$. The distribution $F(\cdot)$, with associated density $f(\cdot)$, $f(\theta) > 0\ \forall\ \theta$ is common knowledge. Each person's θ is her private information.

Our model differs from Mirrlees' in the treatment of the production side of the economy. While we retain the assumption that one unit of effort, n, produces one unit of output, z, measured in units of consumption, we assume that opening a vacancy is costly for the firm. More precisely the firm must pay a cost $\kappa > 0$, also measured in units of consumption, for each job vacancy it creates.

A job opportunity is a contract specifying an output, z, to be produced and a payment y to which the worker is entitled upon producing z. We endow each firm with the ability to commit to a contract offer.[4] When deciding on which and how many work contracts to offer each firm takes as given, for each type of contract the ratio λ of workers applying for the same job and vacancies.

A worker who applies to a given job has a probability p of finding a job, where p is a function of λ. Intuitively, one can imagine that it is not too hard for a worker to observe a contract offer. But, without a centralized mechanism coordinating search efforts, noth-

[4]There is no renegotiation after a match is realized, which means that the wage compression effect highlighted in Hungerbühler et al. (2006), Lehmann et al. (2016) does not play a role here.

ing prevents multiple workers to reach for the same offers, if all other workers are able to observe the same offer. Such behavior leads some workers to remain unemployed.

In most of what follows, with some abuse in notation, we consider the inverse function $\lambda = \lambda(p)$ which maps the number of vacancies per worker which is necessary for the probability that a job applicant finds a job is p. We assume that $\vartheta(p) := \lambda\left(p\right)^{-1}$ is a strictly convex, strictly increasing and twice differentiable function satisfying $\vartheta(0) = 0, \lim_{p\uparrow 1}\vartheta(p) = \infty$.

The economy is also inhabited by a benevolent planner/government who designs redistributive and insurance policies to maximize a social objective to be specified. We assume that the number of vacancies and the type of contract, (z, y), being offered by each firm is publicly observed. This allows one to recover the probability, p, of actually lending a job (z, y), for each (z, y). The planner observes the actual earnings, y, of all employed agents, but not θ. In Section 3, when we take on the issue of implementation, we discuss the importance of observing these variables.

2 Constrained Efficient Allocations

In this section we derive, in the spirit of Werning (2007), general features that any constrained efficient allocation must possess. Contrary to him we focus on allocations, and not tax schedules since more than one policy instrument must be used to implement these allocations.

Planner's Problem To characterize the constrained efficient allocation we consider a mechanism under which each agent announces his type, θ, and is assigned: a specific firm/market to which apply and how much to produce, $z(\theta)$, in case he receives a job offer. Associated with each firm/market is a probability of receiving a job offer, $p(\theta)$, a consumption level $c(\theta)$ conditional on working (and producing $z(\theta)$) and an unemployment benefit $c^u(\theta)$.

As it turns, it is simpler to work with the transformations: $u(\theta) := \varphi(c(\theta))$, $h(\theta) := \eta(z(\theta))$. Define $C(\varphi(c)) := c$, $N(\eta(z)) := z$, and $\underline{u}(\theta) = \varphi(c^u(\theta))$. In this case, we define an allocation as a mapping $\Theta \mapsto [0,1] \times \mathbb{R} \times \mathbb{R}_+ \times \mathbb{R}$, which associates to each type θ a bundle $(u(\theta), \underline{u}(\theta), h(\theta), p(\theta))$.

An allocation is *incentive-feasible* if it satisfies the resource constraint,

$$\int \left\{ p(\theta)\left[N\left(h(\theta)\right) - C(u(\theta))\right] - (1-p(\theta))C(\underline{u}(\theta)) - \frac{\kappa}{\lambda\left(p(\theta)\right)} \right\} f(\theta)d\theta \geq G,$$

for some exogenous G, and the incentive compatibility constraint,

$$\theta \in \underset{s\in\Theta}{\operatorname{argmax}}\, p(s)[u(s) - \theta h(s)] + [1 - p(s)]\underline{u}(s) \quad \forall\theta.$$

An incentive feasible allocation is *efficient* if there is no other incentive feasible allocation which delivers the same expected utility $w(\theta)$ for all θ using fewer resources.[5]

If we define

$$w(\theta) := p(\theta)\,(u(\theta) - \theta h(\theta)) + (1 - p(\theta))\underline{u}(\theta),$$

incentive compatibility is equivalent to the envelope,

$$\dot{w}(\theta) = -p(\theta)h(\theta), \quad \forall\theta$$

and the monotonicity condition,

$$p(\theta)h(\theta) \text{ decreasing.}$$

Notice that the monotonicity constraint guarantees that the second order condition is also satisfied at any allocation satisfying the first order condition of the worker's optimization problem. It is also possible to show that the first and second order conditions guarantee not only a local but a global maximum as in Mirrlees (1971). A novel feature of the problem studied here is that the second order condition entails monotonicity on $p(\theta)h(\theta)$ not on $p(\theta)$ and $h(\theta)$ separately.

Moral Hazard It is not hard to show that any constrained efficient allocation displays full insurance: $c(\theta) = c^u(\theta)$ for all θ. The reason for such counterfactual finding is that we have assumed away a moral hazard problem that unemployment insurance generates. Indeed, to implement a full insurance allocation the planner must not only be able to observe whether an agent actually applied to a specific type of vacancy but also be assured that he does accept any job offers he receives.

If, instead, the planner only observes the vacancy to which an agent applies when the agent actually gets the job, then full insurance cannot be implemented. An agent can always claim to have applied to a vacancy $(z(\theta), y(\theta)$ and/or actually apply to such a job and reject all job offers he receives. By doing so, the agent guarantees himself a utility $\underline{u}(\theta)$ without working. Since this is true for any θ, the only incentive compatible

[5]It is not hard to see that surpluses can be returned to agents in such a way as to increase everyone's utility in an incentive compatible way.

announcement is to claim that one is of type $\underline{\theta}$ but has not found a job.

To avoid such deviation, we add the constraint,

$$w(\theta) \geq \sup_{\tilde{\theta}} \underline{u}(\tilde{\theta}).$$

to the planner's program. This leads to Claim 2.1, below.

Claim 2.1. *If applications are only observed (by the planner) for those who actually get a job, then in any solution to the planner's problem we have $\underline{u}(\theta) = \underline{u}$ for almost every θ.*

2.1 Necessary Conditions

The main goals of this paper are to provide means for identifying the presence of inefficiency in observed allocation and to try to quantify the welfare costs of such inefficiencies. In this section we address the first objective by offering a partial characterization of constrained efficient allocations. This provides us to simple ways of ruling out efficiency of observed allocations. Because the conditions we derive are only necessary, the fact that an allocation display the properties uncovered does not rule out inefficiency.

Our first remark is that since $w(\theta)$ is decreasing in θ, a necessary and sufficient condition for an allocation to be implementable is $w(\bar{\theta}) \geq \sup_{\tilde{\theta}} \underline{u}(\tilde{\theta})$.

Due to Claim 2.1, we know that the unemployment benefit must be independent of θ at the solution of the planner's. Let $\underline{u}$ denote the utility from consumption associated with this type-independent unemployment benefit. Under this moral hazard restriction we define an allocation simply as a mapping $\Theta \mapsto [0,1] \times \mathbb{R} \times \mathbb{R}_+$ from types to vectors (p, u, h).

For any allocation, $(p(\theta), u(\theta), h(\theta))_\theta$, the expect utility of a type θ worker is

$$w(\theta) = p(\theta)\,[u(\theta) - \theta h(\theta)] + \big(1 - p(\theta)\big)\,\underline{u}.$$

Next, we show that an efficient allocation must be such that the least productive (highest θ) agent for whom $h > 0$, must be indifferent between working and remaining unemployed.

Proposition 1. *If $\hat{\theta}$ is the largest type such that $h(\theta) > 0$, then any efficient allocation must be such that*

$$\underline{u} = u(\hat{\theta}) - \hat{\theta} h(\hat{\theta}).$$

To further characterize constrained efficiency we note that an allocation which delivers an expected utility profile $(\varpi(\theta))_\theta$ is constrained efficient if and only if it solves a dual program of the form

$$\max \int \left\{ p(\theta) \left[N(h(\theta)) - C\left(\frac{w(\theta) - \underline{u}}{p(\theta)} + \theta h(\theta) + \underline{u} \right) \right] \right.$$
$$\left. - (1 - p(\theta)) C(\underline{u}) - \frac{\kappa}{\lambda\left(p(\theta)\right)} \right\} f(\theta) d\theta,$$

subject to

$$\dot{w}(\theta) = -p(\theta) h(\theta), \tag{2.1}$$

$$p(\theta) h(\theta) \text{ decreasing,} \tag{2.2}$$

and, $\forall \theta$

$$w(\theta) \geq \varpi(\theta). \tag{2.3}$$

The planner's dual problem, which we denote $\mathscr{P}^{EF}$, is an optimal control program with controls $h(\theta)$ and $p(\theta)$ and state $w(\theta)$. Note that it is an optimal control problem with inequality constraints on the state, $w(\theta)$. We restrict our attention to $\mathscr{C}^2$ solutions which satisfy the monotonicity condition, and ignore the possibility of bunching.[6]

Whereas in Mirrlees (1971), incentive compatibility requires monotonicity in $h(\theta)$, hence in $u(\theta)$, here an implementable allocation must be monotonic in $p(\theta)h(\theta)$, not necessarily in each one separately. Incentive compatibility alone does not guarantee monotonicity. The next proposition shows that *efficiency* implies monotonicity on both $p(\theta)$ and $h(\theta)$.

Proposition 2. *If an allocation is efficient then both $h(\theta)$ and $p(\theta)$ are decreasing in θ.*

Since incentive compatibility only requires monotonicity in $p(\theta)h(\theta)$ while efficiency requires monotonicity in each one separately, observing an allocation that violates monotonicity indicates inefficiency. The following is an example of such an allocation.

Example 2.1. *Consider an economy in which a type θ worker/consumer has preferences of the form $U(c,n) = \ln c - \theta n^2/2$. Assume also that in this economy labor income taxes*

[6]The presence of inequality constraints on state variables poses no additional complexity to our approach provided that we restrict our attention to smooth allocations. See Seiestad and Sydsæter (1987), ch. 5.

are of the form $T(y) = y - \xi y^{1-\tau}$. Finally, let $\lambda(p) = (1/p-1)^{-1}$. In this case, it is not hard to see that

$$z = \frac{1}{2\theta}\left\{\frac{\kappa}{1-p} + \sqrt{\left[\frac{\kappa}{1-p}\right]^2 + 4\theta(1-\tau)}\right\},$$

which allows us to easily find $p(\theta)$ numerically – see Figure 1.

As one can see, for very low values of θ (very high productivity), $p(\theta)$ is increasing in θ – see Figure 1. According to Proposition 2, the equilibrium allocation is inefficient.

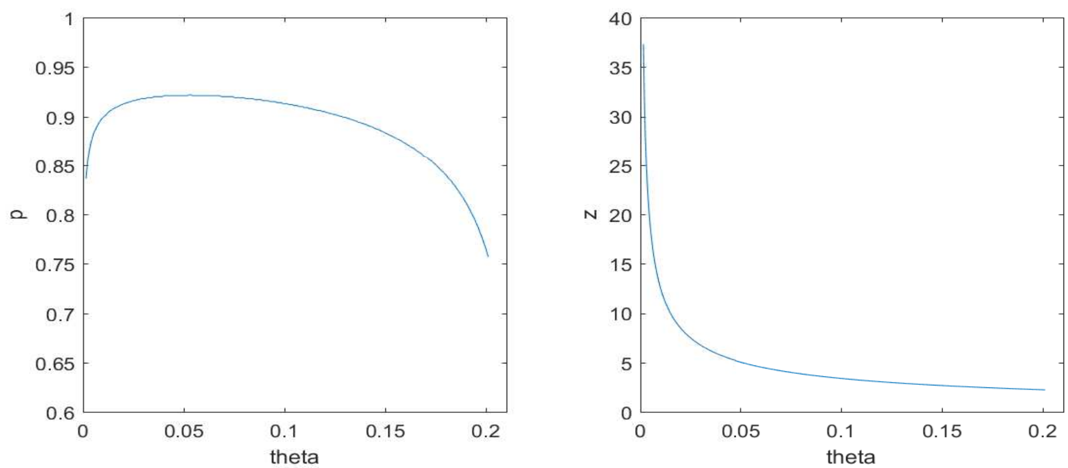

Figure 1: **Inefficient Equilibrium** The two panels display equilibrium $p(\theta)$ and $z(\theta)$ for an economy with $\lambda(p) = p/(1-p)$, $T(y) = y - \xi y^{1-\tau}$, $\varphi(c) = \ln c$, $\eta(z) = z^2/2$. We have used $\tau = 0.05$, $\xi = 0.8$ and $\kappa = 0.005$.

The sign of optimal wedges depends on the objective. The well known findings regarding the sign of wedges, e.g., Seade (1982), concern a specific point in the Pareto frontier. In Mirrlees' (1971) original paper, the objective is Utilitarian, with positive wedges proved to be optimal also in the Rawlsian case, e.g., Phelps (1973); Boadway and Jacquet (2008).[7] Unless we specify the society's goal we will not be able to pin down the sign of wedges. Yet, because two different margins are defined here, effort and vacancy margins, we are able to relate them for any constrained efficient allocation.

Associate to the two margins the *effort wedge*,

$$\tau^n(\theta) := N'(h(\theta)) - \theta C'(u(\theta)),$$

[7]For the separable case, an early proof is in Mirrlees (1971). Seade (1982) assumes away bunching, while Ebert (1992) handles the general case. Werning (2000) has a simpler, more direct proof.

and the *vacancy wedge*,

$$\tau^p(\theta) := N(h(\theta)) - C(u(\theta)) + C(\underline{u}) + C'(u(\theta))\frac{w(\theta) - \underline{u}}{p(\theta)} - \kappa\vartheta'(p(\theta)).$$

The next proposition shows that the effort and the vacancy wedges have the same sign *at all productivity levels* and *for any Pareto efficient allocation.*

Proposition 3. *At any constrained efficient allocation, agents facing non-negative (non-positive) wedges on effort face non-negative (non-positive) wedges on vacancy offers too.*

Proposition 3 provides a simple efficiency test: any positive (negative) marginal taxes on labor income, must be accompanied by positive (negative) taxes on vacancy offers.

The role of market frictions An important general finding in the optimal tax literature that applies to most efficient allocations in a pure Mirrlees setting relates to the taxation of least productive agents. When inequality aversion is finite and non-negative marginal tax rate is zero at the bottom, unless there is bunching – Seade (1977). The presence of labor market frictions overturns this result; the next proposition shows that marginal tax rates are positive at $\bar{\theta}$ for any allocation which maximizes a weighted Utilitarian objective.

Proposition 4. *If an allocation maximizes a weighted Utilitarian objective,*

$$\int \alpha(\theta) w(\theta) dF(\theta),$$

$\alpha(\theta) \geq 0$, $\int \alpha(\theta) dF(\theta) = 1$, *then:*

a - the effort and vacancy wedges faced by the least productive worker are positive, and;

b - both wedges faced by the most productive worker are zero.

Proposition 4 does not restrict the set of weights except by non-negativity, which is required for Pareto Efficiency. This means that it covers a large subset of all constrained efficient allocations. In fact, for a convex set of attainable utilities, any constrained efficient allocations can be derived as the solution of a weighted utilitarian program.

The typical argument for zero taxes at the bottom is as follows. Consider the impact of slightly increasing marginal tax rates at an income level z'. This introduces a distortion, hence a welfare loss, in the allocation of agents at this income level. On the other hand, the extra revenue raised from agents earning more than z', allows one to spare agents earning z' or less from paying more taxes. Hence provided that society weights the difference between total welfare gains for difference in social value of total welfare gain for those below z' and losses for those above z' less than the welfare cost imposed on agents exactly at z', this tax increase is worth making. At the bottom, the mass of agents at or below z' is by definition zero, which, due to finite inequality aversion, means zero social weight. Hence, there is no gain to offset the welfare loss. Here, revenues raised from an increase in marginal tax rate means that we need not reduce the unemployment benefit which has a welfare value that is always positive for it applies to the whole distribution of agents. Hence, positive marginal tax rates remain optimal even at the bottom.

This logic exposed in the previous paragraph does not apply to the Rawlsian (infinite inequality aversion) case, of course, where full weight is assigned to the measure zero set of agents at the bottom of the productivity distribution. That is, for a Rawlsian metric the optimum in a Mirrlees' setting displays positive taxes at the bottom. Proposition 5 shows that another property of optimal taxes remains true in our setting: taxes are everywhere non-negative.

To show this, we first note that an allocation solves the Rawlsian program if it solves the cost minimization program when constraint (2.3) is replaced by

$$w(\bar{\theta}) \geq \varpi(\bar{\theta}). \tag{2.4}$$

The next proposition extends Mirrlees' (1971) findings to our setting.

Proposition 5. *An allocation that solves a Rawlsian objective exhibits positive effort and vacancy wedges at all productivity levels except the very top.*

In a Mirrlees's setting very detailed characterizations are available for the Rawlsian objective. In particular, the relationship between the underlying distribution of productivities and the properties of optimal tax schedules have been provided by Kanbur and Tuomala (1994); Boadway and Jacquet (2008); Hellwig (2010). Moreover, the Rawlsian case corresponds to a revenue maximization objective – Piketty (1997) – which is associated with the maximum size of the Government. Finally,

3 Implementation via labor income taxes?

Assume that the planner uses a labor income tax schedule and a type independent unemployment benefit as instruments. In this section we ask if these instruments suffice to implement the constrained efficient allocations from Section 2.

Let $T(\cdot)$, a labor income tax, be in place. Implicitly define the function $\chi : \mathbb{U} \mapsto R$ through, $\varphi\left(\chi(u) - T(\chi(u))\right) = u$ for all $u \in \mathbb{U}$.

In this case, a firm which decides to offer type θ workers employment contracts (u, h) solves a problem of the form

$$\min_{u,h,p} \lambda(p) p \left[\chi(u) - N(h)\right] + \kappa$$

s.t.

$$p\left[u - \theta h - \underline{u}\right] + \underline{u} \geq w(\theta). \tag{3.1}$$

The firm's optimization problem is, therefore, very similar to the autarky program. The difference is only that $\chi(\cdot)$ substitutes for $C(\cdot)$ as the cost function associated with the utility from consumption u. $\chi(\cdot)$ differs from $C(\cdot)$ because it takes into account the labor income tax schedule set in place by the government.

Also note that no incentive constraints are imposed on the firm. In fact, this is a private values environment in a competitive setting – see Pouyet et al. (2008). The firm's profit is independent of a worker's type, *conditional on a labor contract* (z, y).

As it turns, to find the equilibrium allocation it is simpler to rely on the fact that any equilibrium allocation solves for every θ the following problem

$$\max_{p,h,u} p\left[u - \theta h - \underline{u}\right]$$

subject to

$$N(h) - \chi(u) \geq \frac{\kappa}{\lambda(p) p},$$

which we denote program $\mathscr{P}^{EQ}$.

We can use the zero profit condition above along with the firms' first order conditions to show that the planner cannot implement the constrained efficient allocation derived in Section 2.

Proposition 6. *The allocation $(p^\star(\theta), h^\star(\theta), u^\star(\theta))_\theta$ which solves the planner's Utilitarian program cannot be implemented using only a non-linear labor income tax schedule.*

To understand the rationale for this finding recall that Claim 2.1 has established that, at the optimum, $u(\bar{\theta}) - \bar{\theta} h(\bar{\theta}) = \underline{u}$. The agent is in this case indifferent between any two values of p.

Since $p(\theta)$ is decreasing in θ, then we know that $p(\bar{\theta}) < 1$. The profit of a firm hiring $\bar{\theta}$ types is

$$\lambda(\bar{\theta}) p(\bar{\theta}) \left[N(h(\bar{\theta})) - \chi(u(\bar{\theta}))\right] - \kappa \geq 0 \Rightarrow N(h(\bar{\theta})) > \chi(u(\bar{\theta}))$$

If the firm offers a slightly higher utility $u(\bar{\theta}) + \epsilon$, such that $N(h(\bar{\theta})) > \chi(u(\bar{\theta}) + \epsilon)$, it will make the agent strictly prefer to be employed and will allow the firm to raise $p(\bar{\theta})$ as much as it wants thus increasing profits.[8]

Observing h Suffices for Implementation Next, we show that if the government observes h, or, equivalently, the agent's output, $N(h)$, then the second-best can always be implemented.

When $h \in \{h(\theta)\}_{\theta \in \Theta}$ is observable we can assume that the firm truthfully reports h to the government, which charges the firm a total payment $Z(\theta)$ implicitly defined by:

$$\lambda\left(p(\theta)\right) p(\theta) \left[N(h(\theta)) - Z(\theta)\right] - \kappa = 0. \tag{3.2}$$

The government, then, makes a transfer equal to $C(u(\theta))$ to a worker who has produced $N(h(\theta))$. Each vacancy is, in this sense, associated with a production level $N(h(\theta))$. By the zero-profit condition above, the number of vacancies per worker of a firm requiring effort $h(\theta)$ is $\lambda\left(p(\theta)\right)^{-1}$.

Clearly, no firm nor worker has any incentive to deviate. Next, using (3.2) one verifies that the government's revenue is:

$$\int \left\{p(\theta)\left[Z(h(\theta)) - C(u(\theta))\right] - (1 - p(\theta))C(\underline{u})\right\} f(\theta) d\theta$$
$$= \int \left\{p(\theta)\left[Z(h(\theta)) - C(u(\theta))\right] - (1 - p(\theta))C(\underline{u}) - \frac{\kappa}{\lambda\left(p(\theta)\right)}\right\} f(\theta) d\theta,$$

which is the value attained in the second best.

As we have seen, a vacancy is a pair (z, y) to which a firm commits for the agent who

[8]Needs only to choose ϵ and $\hat{p} < 1$ in such a way that

$$N(h(\bar{\theta})) - \chi(u(\bar{\theta}) + \epsilon) > \frac{\lambda(p(\bar{\theta})) p(\bar{\theta})}{\lambda(\hat{p})\hat{p}} \left[N(h(\bar{\theta})) - \chi(u(\bar{\theta}))\right].$$

first arrives at her door. Output, $z = N(h)$, is therefore observable and the allocations derived from the planner's program in Section 2, implementable. In practice, output data may not be so easily available, which begs the question of how much is lost by relying on worker's earnings information only. Hence, in the following section we compute the revenue losses in the current US system, when compared to the revenues raised under an alternative system which delivered the same utility profile while satisfying the first order conditions for the planner's program.

4 Assessing observed allocations

We have shown in the previous section that a labor income tax accompanied only by an unemployment benefit cannot implement all constrained efficient allocations. In this section we derive an alternative allocation which generates the same utility profile as the allocation induced by the US tax system at the same time that it satisfies the necessary conditions for an optimum. We compare the revenues generated by the two allocations. As we shall see, the US current system is sub-optimal.

We start with the status quo labor income tax schedule, $\mathcal{T}(\cdot)$ and let $(\varpi_{\mathcal{T}}(\theta))_\theta$ denote the utility attained by agents under such schedule. Then, we consider program, which we denote $\mathscr{P}^{EF}$ for

$$w(\theta) \geq \varpi_{\mathcal{T}}(\theta), \quad \forall \theta. \tag{4.1}$$

We compare the revenue raised under $\mathscr{P}^{EF}$ with the revenue raised under the status quo schedule. If the former is greater than the latter we can find an allocation which Pareto dominates the status quo.

$\mathscr{P}^{EF}$ is analogous to the program solved by Werning (2007), to assess the constrained efficiency of real world tax schedules. He uses the first order necessary conditions for an optimum to check whether the observed schedules are rationalized by any optimization program leading to Pareto efficient allocations. We, in contrast, use the empirical utility profile $(\varpi_{\mathcal{T}}(\theta))_\theta$ and check whether it satisfies the first order conditions of program $\mathscr{P}^{EF}$, as we shall describe momentarily.

Before, however, it is important to highlight some other differences between Werning' (2007) procedure and ours. First, the crucial element of our analysis is the empirical utility profile $(\varpi_{\mathcal{T}}(\theta))_\theta$, whereas in Werning (2007) it is the tax schedule. Provided that we can assess these equilibrium utilities, our task is rather simple. The procedure for retrieving the utility profile from the data may however be based on assumptions which are not compatible with efficiency, e.g., equilibrium under an incomplete set of policy instruments. In this case it is the quantitative importance not the presence of inefficiency which is of concern. Second, in practice we do not impose (2.2) when we derive an allocation satisfying the first order conditions for $\mathscr{P}^{EF}$. We must check ex-post whether it satisfies (2.2), whereas the use of empirical allocations guarantee incentive compatibility, hence (2.2) in Werning (2007). Finally, program $\mathscr{P}^{EF}$ is not concave. So the first order conditions are only necessary for an optimum, which is in contrast with the Mirrlees' program evaluated by Werning (2007).

Combining (A.5) and (A.6), we have, after some substitutions,

$$N(h(\theta)) - C(u(\theta)) + C(\underline{u}) + C'(u(\theta))\left[u(\theta) - \underline{u}\right] - \kappa\vartheta'(p(\theta)) = N'(h(\theta))h(\theta) \tag{4.2}$$

Now, using $u(\theta) = \theta h(\theta) - \underline{u} + [\varpi_{\mathscr{T}}(\theta) - \underline{u}]/p(\theta)$, and $\dot{\varpi}_{\mathscr{T}}(\theta) = -p(\theta)h(\theta)$, equation (4.2), gives us $h(\theta)$ as a function of $\varpi_{\mathscr{T}}(\theta)$, $\dot{\varpi}_{\mathscr{T}}(\theta)$ and $\underline{u}$. That is, for any $\underline{u}$ and a given path for $\varpi_{\mathscr{T}}(\theta)$, equation (4.2) is a function of $h(\theta)$ only: for a given $\underline{u}$ we use (4.2) to solve for $h(\theta)$ as a function of $\varpi_{\mathscr{T}}(\theta)$ (and $\dot{\varpi}_{\mathscr{T}}(\theta)$), which we recover from the data. Finally, note that $\underline{u}$ is itself a policy choice that must satisfy $\underline{u} = u(\bar{\theta}) - \hat{\theta}h(\bar{\theta})$ if at the optimum all types work with positive probability.

Of course we could have used the same equations to solve for $p(\theta)$ and $u(\theta)$, instead. Hence, provided that we observe $\varpi_{\mathscr{T}}(\theta)$, we can recover the optimum allocation as a function of $\underline{u}$. Hence, all we need to implement the assessment above is to have a procedure to extract from the data, for all θ, the equilibrium utility, $\varpi_{\mathscr{T}}(\theta)$.

4.1 Quantitative Assessment

Assume that we observe $\underline{u}$ and $(p(\theta), h(\theta), u(\theta))_\theta$, then we are, for all purposes, observing the utility profile $(\varpi(\theta))_\theta$. Knowledge of $f(\cdot)$, $\lambda(\cdot)$, and κ, allows us to recover the economy's resource use. Next we can implement $\mathscr{P}^{EF}$ using $(\varpi(\theta))_\theta$ and compare the revenues raised by this program with the revenues raised by the current US policy.

Whether we can and under what assumptions we can recover $(p(\theta), h(\theta), u(\theta))_\theta$ defines the type of question we are able to answer in our quantitative assessment.

We shall first assume that only the distribution of earnings is observed. Moreover, the only policy instruments are an income tax schedule, $T(\cdot)$ and the unemployment benefit, $\underline{u}$. In this case, after making some parametric assumptions, we solve for the equilibrium of the economy. Note that by restricting the instruments we are in practice ruling out by fiat the efficiency of observed allocations.

We assume $\varphi(c) = c^{1-\sigma}/(1-\sigma)$, $\eta(n) = n^{1+\gamma}/(1+\gamma)$ and $p = (1+\lambda)^{-1}$, which implies

$\lambda(p) = 1/p - 1$. Moreover, to parametrize the US tax system, we rely on the functional form $T(y) = y - \xi y^{1-\tau}$ proposed by Musgrave (1959); Feldstein (1969) – and used by Bénabou (2000, 2002); Heathcote et al. (2017), to name a few.

As we have seen in Section 3, page 13, this characterization is easier if we rely on program $\mathscr{P}^{EQ}$. That is, an equilibrium allocation solves, for a firm participating in the θ-market the dual program,

$$\max_{p,y,z} p\left[\varphi(y - T(y)) - \theta\eta(z) - \underline{u}\right]$$

subject to

$$z - y \geq \frac{\kappa}{\lambda(p)p}.$$

4.2 Numeric results

Solving program $\mathscr{P}^{EQ}$ for the functional forms we are using here, one finds

$$-\frac{\bar{u}}{A(y)} + \frac{1-(1-\tau)(1-\sigma)}{(1-\tau)(1-\sigma)}y - \frac{y^2}{\kappa} = -\left[\frac{\gamma}{1+\gamma}A(y)^{\frac{1}{\gamma}} + \frac{2A(y)^{\frac{1}{\gamma}}y}{\kappa}\right]\theta^{-\frac{1}{\gamma}} + \frac{A(y)^{\frac{2}{\gamma}}}{\kappa}\theta^{-\frac{2}{\gamma}}, \quad (4.3)$$

where $A(y) := (1-\tau)\zeta^{1-\sigma}y^{(1-\tau)(1-\sigma)-1}$.

Equation (4.3) provides an analytic expression for θ as a function of $\underline{u}$, κ, and y. It is a maintained assumption of this work that earnings are observed. If we knew κ, $\underline{u}$ and p we would be able to recover, for each y, the preference parameter θ compatible with such choice.

Explain calibration here!!!

The problem is that we do not observe κ, $\underline{u}$ or $p(\theta)$.

Of course if in the data low productivity agents were non-participants, i.e., if there were a mass of agents with such a low productivity that they were never working then, we could use the lowest skill for which agents participate, say θ_0, to pin down $\underline{u}$ through $u(\theta_0) - \theta_0 h(\theta_0) = \underline{u}$.

If we do not observe $p(\theta)$ for each θ, but only on average, we can pick κ to match this average participation rate. If, however, all productivity types participate, then κ and $\underline{u}$ must be simultaneously determined by combining these two expressions.

Insert Sensitivity Exercises

5 Conclusion

TO BE DONE

References

Bénabou, R. (2000). Unequal societies: Income distribution and the social contract. *American Economic Review 90*(1), 96–129. 3, 17

Bénabou, R. (2002). Tax and education policy in a heterogeneous-agent economy: What levels of redistribution maximize growth and efficiency? *Econometrica 70*(2), 481 – 517. 3, 17

Boadway, R. and K. Cuff (2016). Optimal unemployment insurance and redistribution. Working Paper 1375, Queen's Economics Department. 4, 5

Boadway, R. and L. Jacquet (2008). Optimal marginal and average income taxation under maximin. *Journal of Economic Theory 143*(425–441). 10, 12

Ebert, U. (1992). A reexamination of the optimal nonlinear income tax. *Journal of Public Economics 49*, 47–73. 10

Feldstein, M. (1969). The effects of taxation on risk taking. *Journal of Political Economy 77*(5), 755 – 764. 17

Geromichalos, A. (2015). Unemployment insurance and optimal taxation in a search model of the labor market. *Review of Economic Dynamics 18*, 365 – 380. 5

Golosov, M., P. Maziero, and G. Menzio (2013). Taxation and redistribution of residual income inequality,. *Journal of Political Economy, 121(6), pp.116-1204. 121*(6), 1160–1204. 2, 4

Heathcote, J., K. Storesletten, and G. Violante (2017). Optimal tax progressivity: An analytical framework. *Quarterly Journal of Economics forthcoming*. 17

Hellwig, M. F. (2010). Incentive problems with unidimensional hidden characteristics: a unified approach. *Econometrica 78*(4), 1201 – 1237. 12

Hungerbühler, M., E. Lehmann, A. Parmentier, and B. Van der Linden (2006). Opitmal redistributive taxation in an equilibrium search model. *Review of Economic Studies 73*, 743 – 767. 5

Jacquet, L., E. Lehmann, and B. Van der Linden (2013). The optimal marginal tax rates with both extensive and intensive responses. *Journal of Economic Theory 148*(5), 1770 – 1805. 4

Jacquet, N. and S. Tan (2012). Wage-vacancy contracts and coordination frictions. *Journal of Economic Theory 147*(3), 1064 – 1104.

Kanbur, R. and M. Tuomala (1994). Inherent inequality and the optimal graduation of optimal taxes. *Scandinavian Journal of Economics 96*(2). 12

Kroft, K., K. Kucko, E. Lehmann, and J. Schmeider (2017, September). Optimal income taxation with unemployment and wage responses: A sufficient statistics approach. mimeo. 4

Kroft, K., K. Kucko, E. Lehmann, and J. Schmieder (2017, September). Optimal income taxation with unemployment and wage responses: A sufficient statistics approach. Mimeo. University of Toronto. 4

Lehmann, E., M. Hungerbühler, A. Parmentier, and B. Van der Linden (2006). Optimal redistributive taxation in a search equilibrium model. *Review of Economics Studies 73*(3), 743–768. 4

Lehmann, E., B. Lucifora, S. Moriconi, and B. Van der Linden (2016). Beyond the labour income tax wedge: The unemployment-reducing effect of tax progressivity. *International Tax and Public Finance 23*(3), 454–489. 4, 5

Lehmann, E., A. Parmentier, and B. Van der Linden (2011). Optimal income taxation with endogenous participation and search unemployment. *Journal of Public Economic 95*(1523 – 1537). 4

Mirrlees, J. A. (1971). An exploration in the theory of optimal income taxation. *Review of Economic Studies 38*, 175–208. 1, 2, 3, 4, 5, 7, 9, 10, 11, 12, 16

Musgrave, R. A. (1959). *The Theory of Public Finance.* New York. 3, 17

Phelps, E. S. (1973). Taxation of income for economic justice. *Quarterly Journal of Economics 87*(3), 331–354. 10

Piketty, T. (1997). La redistribuition fiscale face au chômage. *Revue française d'économie 12*(1), 175–208. 12

Pissarides, C. A. (1985). Taxes, subsidies and equilibrium unemployment. *Review of Economic Studies 52*(1), 121 – 133. 4

Pouyet, J., B. Salanié, and F. Salanié (2008). On competitive equilibria with asymmetric information. *The B.E. Journal of Theoretical Economics 8*(1 (Topics)), Article 13. 13

Schaal, E. and M. Taschereau-Dumuouchel (2014). Optimal redistributive policy in a labor market with search and endogenous participation. mimeo. 4

Seade, J. (1977). On the shape of optimal tax schedules. *Journal of Public Economics 7*, 203–236. 2, 11

Seade, J. (1982). On the sign of the optimum marginal income tax. *Review of Economic Studies 49*, 637–643. 10

Seiestad, A. and K. Sydsæter (1987). *Optimal Control Theory with Economic Applications* (3rd. ed.), Volume 24 of *Advanced Textbooks in Economics*. North Holland. 9

Werning, I. (2000, June). An elementary proof of positive optimal marginal taxes. Mimeo. University of Chicago. 10

Werning, I. (2007). Pareto efficient income taxation. MIT working paper. 6, 15, 16

A Proofs

Proof of Claim 2.1 Let $\underline{u} := \sup_{\tilde{\theta}} \underline{u}(\tilde{\theta})$ and suppose towards a contradiction that we can find θ such that

$$\underline{u}(\theta) < \underline{u}.$$

For all θ,

$$p(\theta)\,(u(\theta) - \theta h(\theta)) + (1 - p(\theta))\underline{u}(\theta) \geq \underline{u},$$

which implies that $\underline{u}(\theta) < \underline{u} < u(\theta)$. Since $C(\cdot)$ is convex one can construct a least costly allocation by decreasing $u(\theta)$ by $\left(1 - p(\theta)\right)\varepsilon$ and increasing $\underline{u}(\theta)$ by $p(\theta)\varepsilon$, for small enough ε. If this change is possible in a positive-measure set of types $\widetilde{\Theta} \subset \Theta$ then it saves a positive amount of resources, which can then by redistributed by increasing the resulting utilities $u^*(\theta)$ and $\underline{u}^*(\theta)$ uniformly by some $\chi > 0$. □

Proof of Proposition 1 Assume that $\underline{u} < \left(u(\hat{\theta}) - \hat{\theta} h(\hat{\theta})\right)$ and notice that this implies that $\hat{\theta} = \bar{\theta}$ and hence since $w(\theta)$ is decreasing, we can find $\varepsilon > 0$ such that $\underline{u} + \varepsilon < u(\theta)$ and $w(\theta) > \underline{u} + \varepsilon$ for every θ. But then, for every θ one can find $\eta > 0$ such that the allocation can be made less costly by decreasing $u(\theta)$ by $\left(1 - p(\theta)\right)\eta$ and increasing $\underline{u}(\theta)$ by $p(\theta)\eta$, a contradiction. □

Proof of Proposition 2 The first order condition with respect to $p(\theta)$ is

$$\Big\{N(h(\theta)) - C(u(\theta)) + C(\underline{u}) + C'(u(\theta))\left[u(\theta) - \theta h(\theta) - \underline{u}\right] \\ - \kappa \vartheta'(p(\theta))\Big\} f(\theta) = \mu(\theta) h(\theta) \quad \text{(A.1)}$$

The first order condition with respect to $h(\theta)$ is

$$\left[N'(h(\theta)) - \theta C'(u(\theta))\right] f(\theta) = \mu(\theta),$$

which, after multiplying by $h(\theta)$, becomes

$$\left[N'(h(\theta))h(\theta) - \theta h(\theta)C'(u(\theta))\right] f(\theta) = \mu(\theta)h(\theta). \tag{A.2}$$

Substituting (A.2) in (A.1), we get

$$N(h(\theta)) - C(u(\theta)) + C(\underline{u}) + C'(u(\theta))\left[u(\theta) - \theta h(\theta) - \underline{u}\right] \\ - \kappa\vartheta'(p(\theta)) = N'(h(\theta))h(\theta) - \theta h(\theta)C'(u(\theta)),$$

which simplifies to

$$N(h(\theta)) - N'(h(\theta))h(\theta) - C(u(\theta)) + C(\underline{u}) + C'(u(\theta))\left[u(\theta) - \underline{u}\right] = \kappa\vartheta'(p(\theta)). \tag{A.3}$$

Differentiating (A.3) with respect to θ yields

$$N'(h(\theta))\dot{h}(\theta) - N'(h(\theta))\dot{h}(\theta) - N''(h(\theta))\dot{h}(\theta)h(\theta) - C'(u(\theta))\dot{u}(\theta) + \\ C'(u(\theta))\dot{u}(\theta) + C''(u(\theta))\left[u(\theta) - \underline{u}\right]\dot{u}(\theta) = \kappa\vartheta''(p(\theta))\dot{p}(\theta),$$

which simplifies to

$$-N''(h(\theta))h(\theta)\dot{h}(\theta) + C''(u(\theta))\left[u(\theta) - \underline{u}\right]\dot{u}(\theta) = \kappa\vartheta''(p(\theta))\dot{p}(\theta).$$

The agent's first order condition,

$$\dot{u}(\theta) = \theta\dot{h}(\theta) - \frac{\dot{p}(\theta)}{p(\theta)}\left[u(\theta) - \theta h(\theta) - \underline{u}\right] \tag{A.4}$$

allows us to write

$$-N''(h(\theta))h(\theta)\dot{h}(\theta) + C''(u(\theta))\left[u(\theta) - \underline{u}\right]\Bigg\{\theta\dot{h}(\theta) - \\ \frac{\dot{p}(\theta)}{p(\theta)}\left[u(\theta) - \theta h(\theta) - \underline{u}\right]\Bigg\} = \kappa\vartheta''(p(\theta))\dot{p}(\theta),$$

which re-arranging yields

$$\left\{C''(u(\theta))\left[u(\theta)-\underline{u}\right]\theta-N''(h(\theta))h(\theta)\right\}\dot{h}(\theta) \\ =\left\{\kappa\vartheta''(p(\theta))p(\theta)+C''(u(\theta))\left[u(\theta)-\underline{u}\right]\left[u(\theta)-\theta h(\theta)-\underline{u}\right]\right\}\frac{\dot{p}(\theta)}{p(\theta)}.$$

$p(\theta)$ and $h(\theta)$ are, therefore, co-monotone.

The agent's second order condition,

$$\dot{p}(\theta)h(\theta)+p(\theta)\dot{h}(\theta)\leq 0,$$

leads to $\dot{p}(\theta)\leq 0$ and $\dot{h}(\theta)\leq 0$, which completes the proof. □

Proof of Proposition 3 First of all note that the sign of the labor wedge, $\tau^n(\theta)$, is pinned down by the sign of $\mu(\theta)$ through the first order condition with respect to $h(\theta)$,

$$N'(h(\theta))-\theta C'(u(\theta))=\frac{\mu(\theta)}{f(\theta)}.$$

The same is true for the vacancy wedge sign, $\tau^p(\theta)$, which is determined by the first order condition with respect to $p(\theta)$,

$$N(h(\theta))-C(u(\theta))+C(\underline{u})+C'(u(\theta))\frac{w(\theta)-\underline{u}}{p(\theta)}-\kappa\vartheta'(p(\theta))=\frac{h(\theta)\mu(\theta)}{f(\theta)}.$$

Hence, the sign of both wedges coincide. □

Proof of Proposition 4 Both effort and vacancy wedges are pinned down by the sign of μ. Hence, to prove (i), it suffices to note that this is a free boundary program which implies $\mu(\underline{\theta})=0$. For (ii) we need to derive the sign of $\mu(\bar{\theta})$. The planner's dual problem may be written as the following optimal control program,

$$\max\int_{\underline{\theta}}^{\overline{\theta}}\left\{p(\theta)\left[N(h(\theta))-C\left(\frac{w(\theta)-\underline{u}}{p(\theta)}+\theta h(\theta)+\underline{u}\right)\right] \\ -(1-p(\theta))C(\underline{u})-\frac{\kappa}{\lambda\left(p(\theta)\right)}\right\}f(\theta)d\theta,$$

subject to

$$\int_{\underline{\theta}}^{\overline{\theta}}\alpha(\theta)w(\theta)f(\theta)d\theta\geq A,$$

$$\dot{w}(\theta) = -p(\theta)h(\theta),$$

$$w(\bar{\theta}) = \bar{u},$$

and

$$p(\theta)h(\theta) \text{ decreasing.}$$

Here, $h(\theta)$ and $p(\theta)$ are the controls and $w(\theta)$ is the state variable.

We will restrict our attention to C^2 solutions which satisfy the monotonicity condition. We can thus write the Lagrangian:

$$V(A,\underline{u}) = \max \int_{\underline{\theta}}^{\overline{\theta}} \left\{ p(\theta)\left[N(h(\theta)) - C\left(\frac{w(\theta)-\underline{u}}{p(\theta)} + \theta h(\theta) + \underline{u}\right)\right]\right.$$
$$\left. - (1-p(\theta))C(\underline{u}) - \kappa\vartheta(p(\theta)) + \psi\left[\alpha(\theta)w(\theta) - A\right]\right\} f(\theta)d\theta - \mu(\theta)p(\theta)h(\theta).$$

Ignoring bunching, the first order conditions are

$$p(\theta)\left[N'(h(\theta)) - \theta C'(u(\theta))\right] f(\theta) = \mu(\theta)p(\theta), \tag{A.5}$$

$$\left[N(h(\theta)) - C(u(\theta)) + C(\underline{u}) + C'(u(\theta))\frac{w(\theta)-\underline{u}}{p(\theta)}\right.$$
$$\left. - \kappa\vartheta'(p(\theta))\right] f(\theta) = \mu(\theta)h(\theta), \tag{A.6}$$

and

$$-\dot{\mu}(\theta) = -C'(u(\theta))f(\theta) + \psi\alpha(\theta)f(\theta). \tag{A.7}$$

To sign the labor wedge in (A.5) we must assess the sign of $\mu(\theta)$. Integrating $-\dot{\mu}(\theta) = -C'(u(\theta))f(\theta) + \psi\alpha(\theta)f(\theta)$ from $\underline{\theta}$ to $\bar{\theta}$, we obtain

$$-\int_{\underline{\theta}}^{\overline{\theta}} \dot{\mu}(\theta)d\theta = \int_{\underline{\theta}}^{\overline{\theta}} \left[\psi\alpha(\theta)f(\theta) - C'(u(\theta))f(\theta)\right] d\theta,$$

or

$$\mu(\bar{\theta}) = -\int_{\underline{\theta}}^{\overline{\theta}} \left[\psi\alpha(\theta)f(\theta) - C'(u(\theta))f(\theta)\right] d\theta + \mu(\underline{\theta}).$$

Since $\mu(\bar{\theta}) = 0$, we get

$$\mu(\bar{\theta}) = -\int_{\underline{\theta}}^{\overline{\theta}} \left[\psi\alpha(\theta) f(\theta) - C'(u(\theta)) f(\theta)\right] d\theta.$$

Recall that, since $w(\bar{\theta}) = \underline{u}$, we do not necessarily have $\mu(\bar{\theta}) = 0$. We would like to know the sign of $\mu(\bar{\theta})$. Notice that the allocation $\left(u^*(\theta) + x, \underline{u} + x, p(\theta), h(\theta)\right)$ is always feasible for $|x| < \varepsilon$ for some $\varepsilon > 0$.

Therefore, we obtain

$$-\frac{\partial V(A, \underline{u})}{\partial A} = \psi = \int_{\underline{\theta}}^{\overline{\theta}} \left[p(\theta)C'(u(\theta)) + (1 - p(\theta))C'(\underline{u})\right] f(\theta) d\theta,$$

which finally implies

$$\begin{aligned}\mu(\bar{\theta}) = -\int_{\underline{\theta}}^{\overline{\theta}} \left[p(\theta)C'(u(\theta)) + (1 - p(\theta))C'(\underline{u})\right] f(\theta) d\theta + \int_{\underline{\theta}}^{\overline{\theta}} C'(u(\theta)) f(\theta) d\theta \\ = \int_{\underline{\theta}}^{\overline{\theta}} (1 - p(\theta)) \left(C'(u(\theta)) - C'(\underline{u})\right) f(\theta) d\theta > 0.\end{aligned}$$

The least productive agent, faces a positive marginal tax rate. □

Proof of Proposition 5 In the Rawlsian program, the co-state equation becomes

$$-\dot{\mu}(\theta) = -C'(u(\theta)) f(\theta) \quad \forall \theta, \tag{A.8}$$

whereas the optimality condition for the free boundary constraint is $\mu(\underline{\theta}) = 0$. Hence, $\mu(\theta) \leq 0 \; \forall \; \theta > \underline{\theta}$. The result is then immediate from the first order condition with respect to $h(\theta)$. □

Proof of Proposition 6 Differentiating the zero profit condition,

$$N(h(\theta)) - \chi(u(\theta)) = \frac{\kappa}{\lambda(p(\theta)) p(\theta)},$$

we get

$$N'(h(\theta)) \dot{h}(\theta) - \chi'(u(\theta)) \dot{u}(\theta) = -\frac{\kappa[\lambda'(p(\theta)) p(\theta) + \lambda(p(\theta))] \dot{p}(\theta)}{[\lambda(p(\theta)) p(\theta)]^2}.$$

Next, using the firms' first order conditions,

$$N'(h(\theta))\left[\dot{h}(\theta)-\frac{\dot{u}(\theta)}{\theta}\right]=-\frac{\kappa\left[\lambda'(p(\theta))p(\theta)+\lambda(p(\theta))\right]\dot{p}(\theta)}{[\lambda(p(\theta))p(\theta)]^2}.$$

Finally, using the agents' envelope and the definition of $w(\theta)$,

$$N'(h(\theta))\left\{\frac{\dot{p}(\theta)}{\theta p(\theta)}\left[u(\theta)-\theta h(\theta)-\underline{u}\right]\right\}=-\frac{\kappa\left[\lambda'(p(\theta))p(\theta)+\lambda(p(\theta))\right]\dot{p}(\theta)}{[\lambda(p(\theta))p(\theta)]^2}.$$

This must hold for all agents.

But, we know from Claim 2.1 that there is at least one type $\hat{\theta}$ such that

$$\underline{u}=\left(u(\hat{\theta})-\hat{\theta}h(\hat{\theta})\right),$$

which implies

$$\frac{\kappa\left[\lambda'(p(\hat{\theta}))p(\hat{\theta})+\lambda(p(\hat{\theta}))\right]\dot{p}(\hat{\theta})}{\left[\lambda(p(\hat{\theta}))p(\hat{\theta})\right]^2}=0.$$

This condition cannot be satisfied for any $p(\theta)<1$. □

B Derivations for Numeric Findings

B.1 Deriving expression (4.3)

Using our preferred parametrization (and returning to the primal variables), it is then the case that an equilibrium allocation solves

$$\max_{p,z,y} p\left[\frac{1}{1-\sigma}(\zeta y^{1-\tau})^{1-\sigma}-\frac{\theta}{1+\gamma}z^{1+\gamma}-\underline{u}\right]$$

subject to

$$z-y\geq\frac{\kappa}{1-p}. \tag{B.1}$$

The first order condition for this problem are

$$p(1-\tau)\zeta^{1-\sigma}y^{(1-\tau)(1-\sigma)-1}=\alpha(\theta), \tag{B.2}$$

with respect to y,

$$\theta p z^{\gamma}=\alpha(\theta), \tag{B.3}$$

with respect to z, and

$$\frac{1}{1-\sigma}(\zeta y^{1-\tau})^{1-\sigma} - \frac{\theta}{1+\gamma} z^{1+\gamma} - \underline{u} = \alpha(\theta)\frac{\kappa}{\left(1-p\right)^2} \tag{B.4}$$

with respect to p.

Using (B.2) and (B.3) we get

$$\left[(1-\tau)\zeta^{1-\sigma} y^{(1-\tau)(1-\sigma)-1}\right]^{\frac{1}{\gamma}} \theta^{\frac{1}{-\gamma}} = z \tag{B.5}$$

which shows that no extra distortion, beyond that caused by labor income taxes, is introduced by the firm.

Multiplying (B.2) by $y/((1-\tau)(1-\sigma))$, (B.3) by $z/(1+\gamma)$, and adding the two we get

$$p\left[\frac{(\zeta y^{1-\tau})^{1-\sigma}}{1-\sigma} - \theta\frac{z^{1+\gamma}}{1+\gamma} - \bar{u}\right] = \frac{\alpha(\theta) y}{(1-\tau)(1-\sigma)} - \frac{\alpha(\theta) z}{1+\gamma} - p\bar{u},$$

which can be written as

$$\alpha(\theta)\frac{p\kappa}{(1-p)^2} = \frac{\alpha(\theta) y}{(1-\tau)(1-\sigma)} - \frac{\alpha(\theta) z}{1+\gamma} - p\bar{u},$$

using (B.4).

Next, using the fact that contraint (B.1) is active at the optimum, we have

$$(z-y)^2 = \frac{\kappa^2}{(1-p)^2}, \quad \text{and} \quad p = 1 - \frac{\kappa}{z-y},$$

which can be used to obtain

$$\bar{u} = \theta z^{\gamma}\left[\frac{y}{(1-\tau)(1-\sigma)} - \frac{z}{1+\gamma} - \frac{(z-y)^2}{\kappa} + (z-y)\right],$$

or

$$\bar{u} = \theta z^{\gamma}\left[\frac{1-(1-\tau)(1-\sigma)}{(1-\tau)(1-\sigma)} y + \frac{\gamma}{1+\gamma} z - \frac{z^2}{\kappa} + \frac{2zy}{\kappa} - \frac{y^2}{\kappa}\right].$$

Let $A(y) := (1-\tau)\zeta^{1-\sigma} y^{(1-\tau)(1-\sigma)-1}$, then we can finally write

$$-\frac{\bar{u}}{A(y)} + \frac{1-(1-\tau)(1-\sigma)}{(1-\tau)(1-\sigma)} y - \frac{y^2}{\kappa} = -\left[\frac{\gamma}{1+\gamma} A(y)^{\frac{1}{\gamma}} + \frac{2A(y)^{\frac{1}{\gamma}} y}{\kappa}\right]\theta^{-\frac{1}{\gamma}} + \frac{A(y)^{\frac{2}{\gamma}}}{\kappa}\theta^{-\frac{2}{\gamma}}.$$

B.2 Observed p

If we observe not only y but also p, then equation (B.2) allows us to define $\tilde{\alpha}(p, y)$ through

$$p(1-\tau)\zeta^{1-\sigma} y^{(1-\tau)(1-\sigma)-1} = \tilde{\alpha}(p, y).$$

Using (B.3) we may re-write (B.4) as

$$z = -\frac{p(1+\gamma)}{\tilde{\alpha}(p, y)} \left[\tilde{\alpha}(p, y) \frac{\kappa}{\left(1-p\right)^2} + \underline{u} - \frac{1}{1-\sigma} (\zeta y^{1-\tau})^{1-\sigma} \right].$$

That is, knowledge of p and y allows us to recover the associated z, hence, θ, through $\theta p z^{\gamma} = \tilde{\alpha}(p, y)$.

Note that, if we define $s(\theta)$ through

$$z(\theta) + s(\theta) - y(\theta) = \frac{\kappa}{1-p},$$

then $s(\theta)$ need not be zero. The use of these resources must be included in the economy's resource constraint.

C Figures

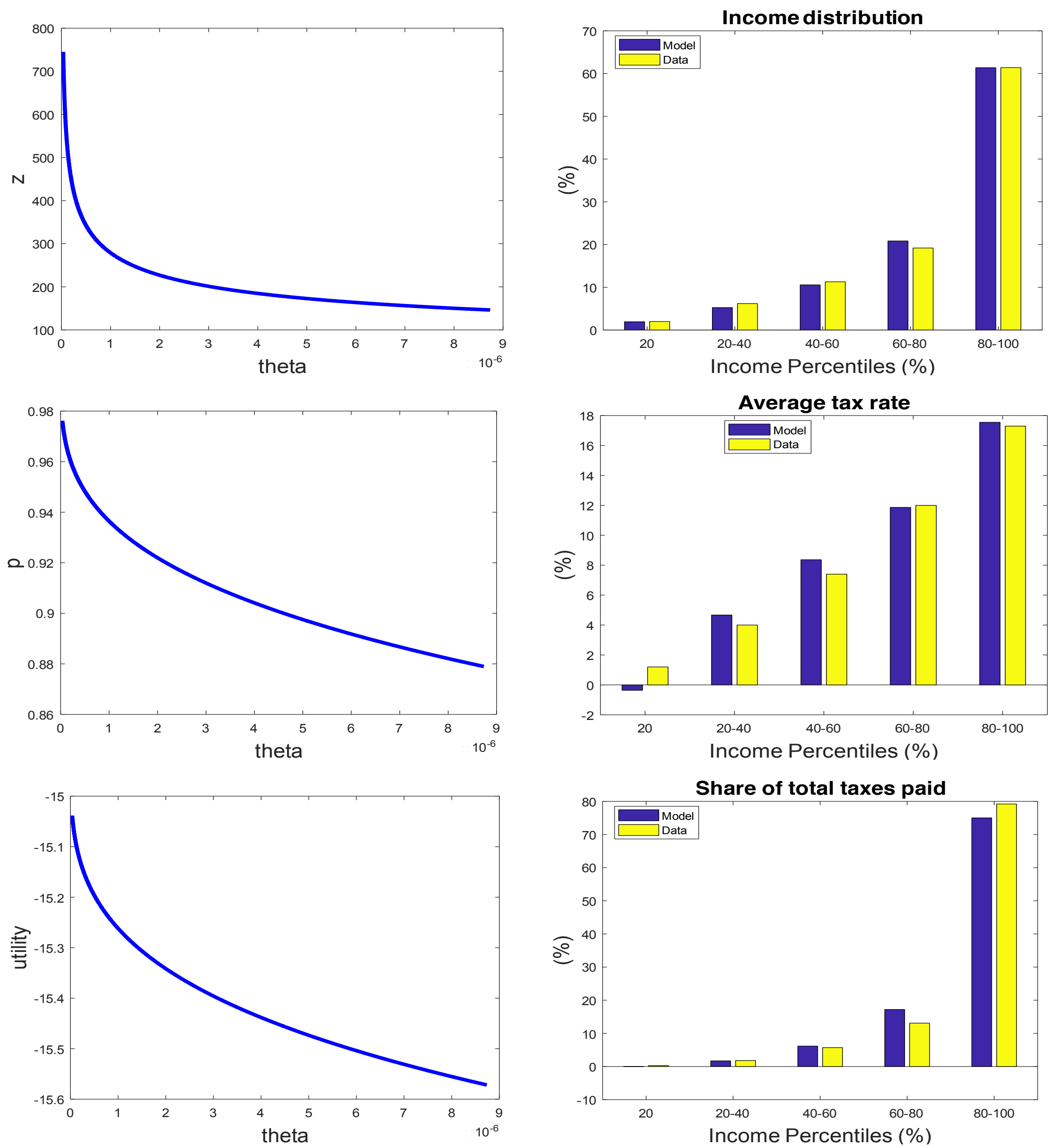

Figure 2: **Calibrated Economy** The panels in the left side of this figure display the behavior of z, p, and w for the calibrated economy. The panels in the right side compare the distribution of income and taxes in the model and in the data.

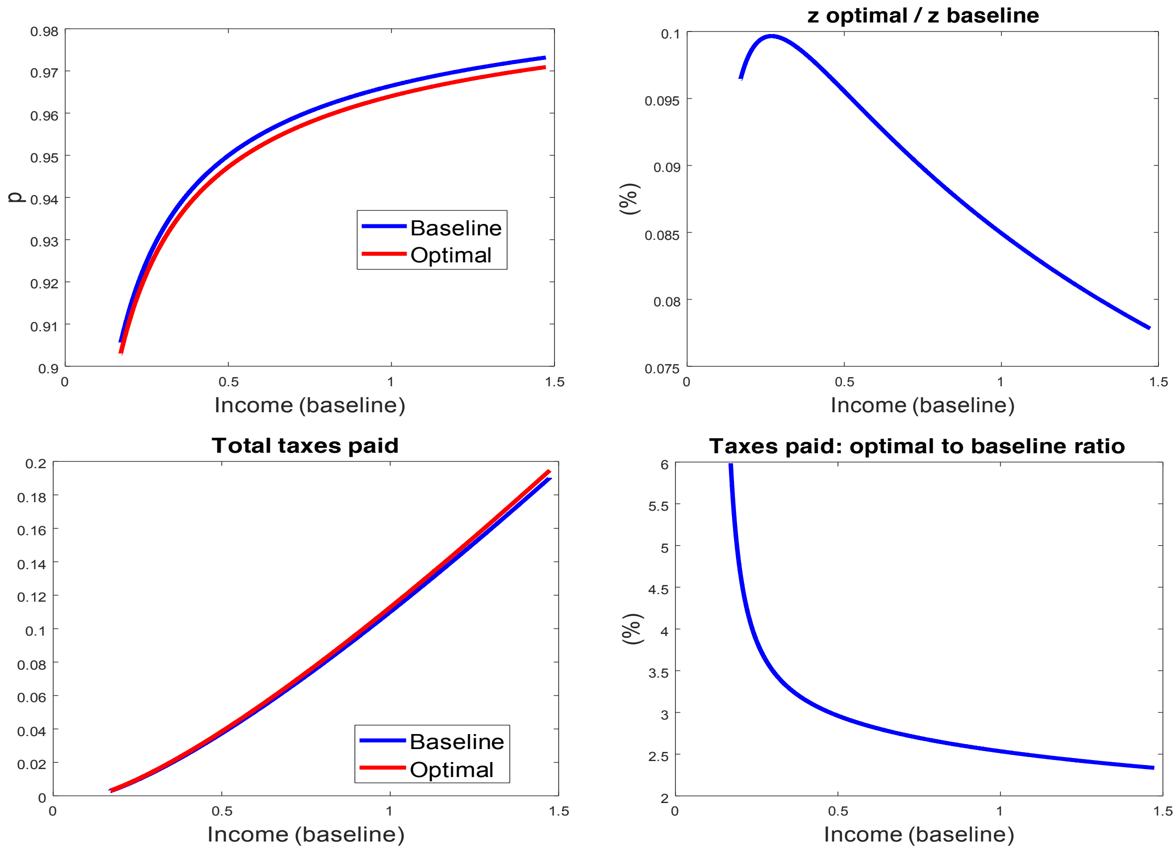

Figure 3: **Optimal Allocation** The panels in the righ display the baseline and the optimal values for p and z. The panels in the right display the ratio between the two.