On Efficient Taxation

Maciej K. Dudek*

May 17, 2013

Abstract

We solve an often neglected problem of efficient taxation where the policy maker is interested in minimizing the dead weight burden imposed on the society by labor income taxes. Specifically, we argue that the efficient tax function assumes the form of a step function. Furthermore, we allow the policy maker to exhibit aversion toward inequality and show that the key result prevails, i.e., that the tax function that ensures efficiency is locally flat.

Key Words: Efficient Taxation, Dead Weight Burden, Step Function, Zero Marginal Taxes, Quasi-Linearity.

JEL: H21

1 Introduction

Taxes at a very fundamental level allow governments to collect revenue, but, typically, create distortions that lead to welfare losses larger than those implied by the tax liabilities themselves. This basic observation leads to a natural, but yet profound, economic problem. How should a policy maker with given revenue needs R decide on the proper form of tax function? This question can be, and has been, answered at many levels. In this paper we propose a novel and at the same time very simple approach to the issue and identify tax functions that solve the optimization problem of the policy maker.

The issue analyzed in the paper is not original. Many researchers have tried to characterize the desired tax function. In fact, there exists an extensive and successful

*The author bears complete responsibility for the content of this paper. No institution should be held responsible for errors that remain. Please send your comments to mkdudek@alumni.caltech.edu.

literature that examines the concept of optimal taxation. Specifically, the shape of the optimal tax function has been of interest to economists for a number of decades now. The contribution of Mirrlees [4] offered enough structure on the optimal tax problem to allow for the derivation of the relevant conditions that define the optimal tax function. Despite the complexity of the efficiency conditions, researchers have been able to resort to numerical techniques to identify the shapes of tax functions that solve the optimal taxation problem as defined by Mirrlees. Furthermore, more recent and ingenious approaches of Diamond [1] and Saez [6] have allowed to relate the underlying optimal taxation problem to observables and generate policy relevant implications.

In this paper, we choose to embark on a path that parallels that typically explored in the literature. Specifically, we focus on tax functions that, in addition to ensuring that the revenue needs are met, lead to the lowest possible level of distortions in the economy, i.e., to tax functions, termed *efficient*, that ensure the highest possible degree of efficiency in the economy. Our approach, thus, is much more modest than, but at the same time can be considered dual to, that normally studied in the literature. We are for the most part interested in efficiency whereas the main stream of the literature takes a broader perspective and focuses on welfare.

At the technical level we choose to, following Saez [7], identify departures from efficiency with dead weight burden. Furthermore, initially, we choose to abstract completely from any equity considerations and focus strictly on efficiency issues. Specifically, we are interested in identifying tax functions that minimize the dead weight burden imposed on the society. The problem of minimizing the dead weight burden has not been subject to extensive research with the exception of the work of Saez [7], who derives an explicit analytic expression that defines the dead weight burden minimization function when lump sum taxes are unavailable and individual productivities are not observable. It appears that the literature has mostly ignored the problem of dead weight burden minimization as the problem in its purest form is trivial, and the solution can be degenerate, and, most importantly, the problem has been superseded by a more general problem of optimal taxation. In this paper, we resurrect the problem of dead weight burden minimization and argue that the problem can be of practical relevance. Furthermore, we show that the tax function that ensures efficiency is of particularly simple form and that it is, contrary to the findings of Saez [7], locally flat implying mostly zero marginal tax rates. In addition, we extend our analysis and allow, as it was originally done by Mirrlees [4], for aversion towards inequality on the part of policy makers. We prove that our underlying result is robust, i.e., the shape of the efficient tax function does not change in the qualitative sense, i.e., it remains locally flat, if one introduces a motive for redistribution.

It may appear unnecessary to even examine the issue of efficient taxation given the success of the optimal taxation program. Naturally, we do not share this view as the problem of efficient taxation can be of interest on its own for a number of reasons. First of all, the problem has not been subject to extensive research apart from the effort by Saez [7]. Moreover, as we argue in the paper the approach to the problem proposed by Saez [7] despite being mathematically sound is too restrictive and fails to identify the tax function of interest properly. Consequently, it appears worthwhile to recast the problem and identify the true efficient tax function.

Secondly, it is in fact the case that the problem of efficient taxation − studied in this paper − is in fact more natural and better grounded in economic theory than the canonical problem of optimal taxation. Observe that relying in the process of evaluating different tax function on the dead weight burden rather than on the levels of utility makes actual comparisons meaningful, and by doing so removes the most critical objection to the theory of optimal taxation. Formally, in our approach, a given tax function is evaluated using a well defined metric expressed in observable units and consistent with individual preferences rather than being judged on a metric based on arbitrary units deprived, according to elementary economic theory, of any economic meaning.

Thirdly, following Mirrlees we allow policy makers to exhibit aversion toward inequality. Specifically, we assume that the preferences of policy makers are defined directly over the magnitudes of dead weight burden borne by individual economic agents. Naturally, such an approach does not suffer from the shortcoming of the standard approach that relies on the notion, occasionally criticized see Stiglitz [11], that individual utilities can be subject to cross personal comparisons. Clearly, by defining the preferences directly over meaningful and comparable concepts we not only extend the literature, but also present a framework that can be thought as deeply grounded in economic theory rather than being contrived to deal with a problem at hand.

We establish our findings in a very simple manner that involves only elementary calculus and very rudimentary distortions of tax functions. Specifically, our technique is very simple and is based on very basic lump-sum distortions coupled with a replacement of a part of tax function with a flat portion. We show that distortions of this form, when applied to the efficient tax function, can be revenue neutral while reducing at the same time the amount of distortions in the economy. Consequently, we argue that a given tax function cannot be efficient unless it is locally flat. Furthermore, at the technical level, we use, following Saez [6] and Diamond [1], a utility function that is essentially of a quasi-linear form. Such a specification is not innocuous and makes our underlying method applicable. Nevertheless, our specific choice of utility function does not, as pointed out by Saez [6], contradict empirical evidence as it appears that actual income effects are small.

The approach to optimal taxation proposed by Mirrlees has rightly gained its prominence in the literature. Consequently, many readers may consider modifications, even those offering new insights and as natural as ours, of the original problem to be unnecessary. Furthermore, numerous researchers may believe that our assumptions essentially allow us, especially given the simplicity of our approach, to trivialize a serious economic problem. We strongly disagree, as we study a legitimate problem, and we obtain our results in a framework that is frequently analyzed in the literature, e.g., Diamond [1]. Furthermore, many researchers including Saez and Stantcheva [8], and Werning [12], have pointed out that the original approach of Mirrlees may be too restrictive. Consequently, they argue, it may be necessary to reformulate the original optimal taxation problem to allow for a richer class of objective functions than the utilitarian criterion − a challenge we accept in this paper. Moreover, our results, at a technical level, do not stand in conflict with the literature on optimal taxation. We in fact consider our findings to be complementary to the results of Diamond [1], Saez [6], and Mirrlees [4], as we effectively enrich the class of problems and the class of solutions to taxation problems.

This paper is divided into five sections. We outline the basic problem in the next section. In the following section we present the solution to a simple problem of dead weight burden minimization. In section four we extend our approach and allow for the presence of a motive for redistribution. Finally, section five contains conclusions.

2 Basic Problem

Following Saez [6] and effectively Diamond [1], let us consider an economy populated with economic agents whose preferences are represented with the following utility function

$$U(c, L) = h(c - \frac{1}{2}L^2), \tag{1}$$

where $h(\ )$ is an increasing and differentiable function. For the most part one can assume as done by Saez [6] and effectively by Diamond [1] that $h(\ )$ is simply a logarithmic function. Note that the utility function is effectively of a quasilinear form, which is, as noted by Saez [6], consistent with observables. Our assumption with regard to the form of the utility function is not innocuous as the functional form makes our underlying method applicable.

Furthermore, let us assume that $[a_L, a_H]$ represents the support of the relevant distribution of skills and let $f(\ )$ denote the corresponding *pdf* of the distribution of skills. Naturally, we assume that an agent whose productivity is equal to a delivers $y = aL$ units of output when she chooses to supply L units of labor.

The problem of an economic agent whose productivity is equal to a and who faces tax function $\tau(y)$ can be summarized as

$$max_{\{y\}} U(c = y - \tau(y), L = \frac{y}{a}) = h(y - \tau(y) - \frac{1}{2}(\frac{y}{a})^2), \tag{2}$$

in which we treat consumption as the numeraire.

The relevant first order condition can be expressed as

$$h^{'}(y - \tau(y) - \frac{1}{2}(\frac{y}{a})^2)(1 - \tau'(y) - \frac{y}{a^2}) = 0, \tag{3}$$

which implicitly defines the optimal[1] income earned, y.

Let us denote the optimal choice of an agent whose productivity is equal to a with y_a. Furthermore, let R_a denote the revenue collected from an agent whose productivity is equal to a, i.e., let $R_a = \tau(y_a)$. Finally, let $U_a = h(y_a - \tau(y_a) - \frac{1}{2}(\frac{y_a}{a})^2)$ denote the realized utility of an agent whose productivity is equal to a.

Before we proceed further we want to acknowledge that in this paper we restrict

[1] Occasionally the first order condition is not sufficient, a fact relevant as pointed out by Lollivier and Rochet[3], and by Ebert [2] in the context of optimal taxation.

attention to tax functions $\tau(\,)$ that satisfy

$$\forall y \in R_+ \cup \{0\} |\ \tau(y) \leq y, \tag{4}$$

i.e., to tax functions that are feasible.

Furthermore, in the main part of the paper, we assume[2] that $0 < a_L$, and that $a_H < \infty$. Note that in this case the revenue collected given tax function $\tau(\,)$ is given by

$$R = \int_{a_L}^{a_H} R_a f(a) da. \tag{5}$$

Moreover, we assume, to exclude the possibility of trivial solutions, that R is big enough, so that a uniform tax of $\frac{R}{\int_{a_L}^{a_H} f(a)da}$ on all agents is not feasible, i.e., we assume that

$$\frac{R}{\int_{a_L}^{a_H} f(a)da} > a_L^2. \tag{6}$$

At this stage we choose to focus on the case when all agents supply a strictly positive amount of labor given $\tau(\,)$. Specifically, we assume that

$$\forall a \in [a_L, a_H]|\ h(-\tau(0)) < h(y_a - \tau(y_a) - \frac{1}{2}(\frac{y_a}{a})^2), \tag{7}$$

where y_a satisfies restriction (3). We consider the alternative case, when some agents choose to be idle in Appendix E.

The concept of dead weight burden is routinely used in applied work. Nevertheless, there are some disagreements with regard to its precise meaning. In this paper, we follow Saez [7] and focus on the definition, which appears to be most commonly accepted in the literature. Specifically, we choose to measure the dead weight burden relying on the notion of equivalent variation, i.e., we define the dead weight burden as the negative of the difference between taxes actually paid by an individual and the lump sum tax that would induce the same level of utility.

The tax function, $\tau()$, is of general form and presumably creates some distortions. Therefore, collecting the specific amount of revenue, $\tau(y_a)$, comes at a cost to the consumer that goes beyond the direct revenue cost. The concept of dead weight burden that we employ here is related to that additional cost. In fact, typically, the consumer could be willing to pay more in taxes to avoid this additional cost. Specifically, if

[2] We discuss the case when $a_H = \infty$ in Appendix D.

lump sum taxes are available, then there are no distortions, assuming that a is large enough, and the level of income earned by a consumer whose productivity is equal to a is given by

$$y_a^F = a^2, \tag{8}$$

which leads to the value of realized utility of

$$U_a^F = h(\frac{1}{2}a^2 - T_a), \tag{9}$$

where T_a denotes the lump sum tax actually paid by the agent.

We can now find the value of T_a, which makes the consumer just indifferent to facing the original tax function, $\tau(\,)$, and paying T_a by equating U_a and U_a^F, which yields

$$h(y_a - \tau(y_a) - \frac{1}{2}(\frac{y_a}{a})^2) = h(\frac{1}{2}a^2 - T_a) \tag{10}$$

and further reduces to

$$T_a = \frac{1}{2}(\frac{a^2 - y_a}{a})^2 + \tau(y_a). \tag{11}$$

Observe that originally the agent actually paid $\tau(y_a)$ in taxes and, at the same time, could be willing to pay up to T_a, given by (11), if lump sum taxes were available. Alternatively, we can say that the discrepancy between T_a and $\tau(y_a)$ reflects the additional cost born by the agent due to the presence of the distortions induced by $\tau(\,)$. We choose to identify this *unnecessary* revenue loss stemming from existing distortions with the dead weight burden. Consequently, we have

$$DWB_a = T_a - \tau(y_a) = \frac{1}{2}(\frac{a^2 - y_a}{a})^2. \tag{12}$$

The level of distortion quantified with expression (12) captures the unnecessary burden borne by economic agents when taxes are distortionary. In our analysis we assume that minimizing this unnecessary burden while ensuring that proper revenue needs are met is the principal motive of the government policy.

3 Dead Weight Burden

We are in a position to define the problem of the government. First of all, we assume that individual characteristics are not publicly observable. Furthermore, we assume that the government is interested in minimizing the economy wide dead weight burden. Therefore, we can state the objective function as

$$min_{\{\tau(\,)\}} DWB = \int_{a_L}^{a_H} DWB_a f(a)da. \tag{13}$$

Note that individual dead weight burdens are expressed in common units. Therefore, we do not encounter standard comparability issues by choosing to express the objective function with equation (13).

Naturally, the government must meet its specific revenue needs, R, and must take into account the fact that economic agents rationally respond to the tax schedule. Formally we state the problem as,

$$min_{\{\tau(\,)\}} DWB = \int_{a_L}^{a_H} \frac{1}{2}(\frac{a^2 - y_a}{a})^2 f(a)da \tag{14}$$

subject to

$$R = \int_{a_L}^{a_H} \tau(y_a) f(a)da \tag{15}$$

and to

$$\forall a \epsilon [a_L, a_H] : \; h^{'}(y_a - \tau(y_a) - \frac{1}{2}(\frac{y_a}{a})^2)(1 - \tau'(y_a) - \frac{y_a}{a^2}) = 0. \tag{16}$$

The above optimization problem, expressed with equations (14), (15), and (16), can be dealt with using the proper mathematical techniques. Let $\tau_E()$ be the solution, possibly obtained with the approach of Saez [7] applied to bounded distributions, of the above problem. Naturally, at this stage we cannot state anything about the qualitative features of $\tau_E(\,)$. However, for purely illustrative purposes let us assume that the actual shape resembles that of the optimal tax function as described by Sadka [5], and Seade [9, 10].

Let y_a^* denote the level of income earned and $L_a^* = \frac{y_a^*}{a}$ the corresponding choice of labor supply of a consumer whose productivity is equal to a and who faces the tax schedule, $\tau_E(\,)$. Observe that the revenue brought by an agent whose productivity is equal to a is given by $R_a^* = \tau_E(y_a^*)$ and that the total amount of revenue collected, naturally, must be equal to R, i.e., we have

$$R = \int_{a_L}^{a_H} \tau_E(y_a^*) f(a)da. \tag{17}$$

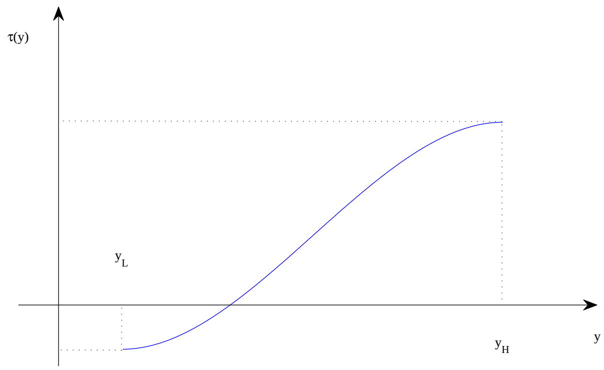

Figure 1: The Generic Form of an Optimal Tax Function.

Furthermore, the realized value of the objective functional is given by

$$DWB^* = \int_{a_L}^{a_H} \frac{1}{2}(\frac{a^2 - y_a^*}{a})^2 f(a)da, \tag{18}$$

and the realized utility of an individual whose productivity is equal to a can be written as

$$U_a^* = h(y_a^* - \tau_E(y_a^*) - \frac{1}{2}(\frac{y_a^*}{a})^2). \tag{19}$$

Recall that we have already assumed that all agents prefer to supply a strictly positive amount of labor, the alternative case is considered in Appendix E, when they face $\tau_E(\ )$, i.e., that

$$\forall a \epsilon [a_L, a_H] | \; y_a^* > 0. \tag{20}$$

Let Y^C denote the set of income levels actually chosen by someone, i.e., let

$$Y^C = \{y_a^* | \; a \in [a_L, a_H]\}. \tag{21}$$

Furthermore, let us now consider a simple variation of $\tau_E(y)$. Specifically, let us

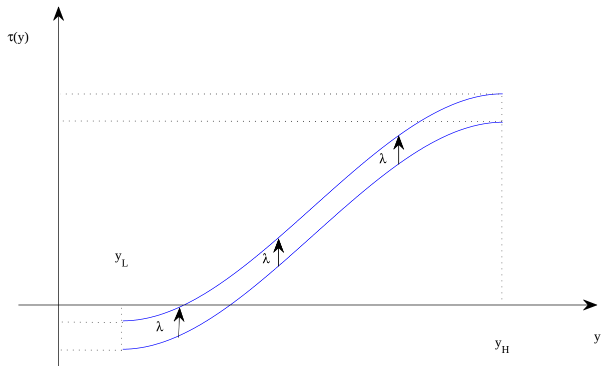

Figure 2: The form of $\tau_\lambda(\,)$ obtained from $\tau_E(\,)$ with an upward shift by λ over the range of actually chosen y.

consider the following tax function

$$\tau_\lambda(y) = \begin{cases} y & \text{if } y \notin Y^C \\ \tau_E(y) + \lambda & \text{if } y \in Y^C \end{cases}, \tag{22}$$

where λ is a, possibly very small, constant.

Figure (2) presents the new tax function, $\tau_\lambda(\,)$. Note that the new tax function given by (22) is simply equal to the previous one shifted up by λ for the relevant range of y. Observe that given our assumptions of strictly positive labor supply on the part of all agents, restriction (7), it must be the case, given the form of the utility function, that the labor supply choice of agents will remain unchanged when tax liabilities are dictated with $\tau_\lambda(\,)$ rather than by $\tau_E(\,)$ for λ sufficiently small[3]. Furthermore, given the form of the utility function and that λ is just a lump sum transfer, there should be no further efficiency losses as compared to the case when tax liabilities are determined with $\tau_E(\,)$. However, the values of individual utilities

[3]Note that $\tau_E(\,)$ is shifted up by λ for all values of y that were originally chosen. Furthermore, for λ small enough it must the case that all agents remain active as originally by assumption they *strictly* preferred to be active.

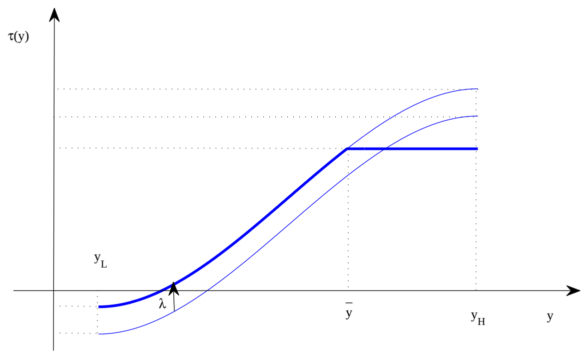

Figure 3: The Form of $\tau_{\bar{y}R}(\,)$.

are affected, and they now become

$$U_a^*(\lambda) = h(y_a^* - \tau_E(y_a^*) - \lambda - \frac{1}{2}(\frac{y_a^*}{a})^2). \tag{23}$$

Similarly, we can, now, express the revenue collected from a single individual as

$$R_a^*(\lambda) = R_a^* + \lambda = \tau_E(y_a^*) + \lambda, \tag{24}$$

which, in particular, implies that the total amount of revenue collected by the government is higher and given by

$$R(\lambda) = \int_{a_L}^{a_H} R_a(\lambda) f(a) da = R + \lambda \int_{a_L}^{a_H} f(a) da, \tag{25}$$

i.e., the government collects more revenue than it needs.

Let us now consider the following class of tax functions

$$\tau_{\bar{y}}(y) = \begin{cases} \tau_\lambda(y) & \text{for } y \leq \bar{y} \\ \tau_\lambda(\bar{y}) & \text{for } \bar{y} < y. \end{cases} \tag{26}$$

Naturally, $\tau_{\bar{y}}(\,)$ looks just like $\tau_\lambda(\,)$, with its right hand end flattened starting from $\bar{y}$, figure (3).

Let us denote with y_L the level of income earned by the lowest type when she faces $\tau_E(\,)$ and y_H denotes the level of income earned by the highest type when taxes are paid according to $\tau_E(\,)$. Furthermore, let us assume that all values of income $y \in [y_L, y_H]$ are actually chosen by someone[4]. Imagine, now, that tax liabilities are to be dictated by $\tau_{\bar{y}}(\,)$ rather than by $\tau_E(\,)$. Observe that when we choose $\bar{y} = y_L$ then the revenue collected from all agents must be, for λ small enough, necessarily smaller than the one that is collected from all agents when taxes are paid in line with $\tau_E(\,)$. Furthermore, when $\bar{y} = y_H$ then $\tau_{\bar{y}}() = \tau_\lambda()$ and the revenue collected from all agents is necessarily equal to the one that is obtained when taxes are paid according to $\tau_\lambda(\,)$ and larger than the revenue collected from all agents when tax liabilities are dictated with the optimal tax function, $\tau_E()$. Therefore, we can expect, relying on an intuitive notion of the mean value theorem[5], that there exists a value of $\bar{y}$ for which he revenue collected from all agents, who face $\tau_{\bar{y}}(\,)$, is the same as the revenue collected from agents when agents face the optimal tax function, $\tau_E(\,)$. Let us denote such a value of $\bar{y}$ with $\bar{y}_R$.

Let us now assume that agents' tax liabilities are determined by $\tau_{\bar{y}_R}(\,)$. Furthermore, let the labor supply choice, given tax function $\tau_{\bar{y}_R}()$, of an agent whose productivity is equal to a be denoted with L_a^{**} and the corresponding income earned with $y_a^{**} = aL_a^{**}$, and let the level of revenue collected be equal to

$$R_a^{**}(\lambda, \bar{y}_R) = \tau_{\bar{y}_R}(y_a^{**}). \tag{27}$$

Note that given the choice of $\bar{y}_R$ the revenue collected with $\tau_{\bar{y}_R}(\,)$ must be the same as the revenue collected with the original optimal tax function, $\tau_E(\,)$, i.e., we must have

$$\int_{a_L}^{a_H} R_a^{**}(\lambda, \bar{y}_R) f(a) da = \int_{a_L}^{a_H} \tau_{\bar{y}_R}(y_a^{**}) f(a) da = \int_{a_L}^{a_H} R_a^* f(a) da = R. \tag{28}$$

Furthermore, the realized utility in this case is given by

$$U_a^{**} = h(y_a^{**} - R_a^{**}(\lambda, \bar{y}_R) - \frac{1}{2}(\frac{y_a^{**}}{a})^2). \tag{29}$$

[4]At this stage, we make this assumption to preserve clarity of the exposition. We realize that this assumption need not be satisfied. We consider the more realistic case in Appendix A.

[5]We provide a rigorous argument in Appendix B.

Finally, the realized value of the objective functional, the dead weight burden, can be expressed as

$$DWB^{**} = \int_{a_L}^{a_H} \frac{1}{2}(\frac{a^2 - y_a^{**}}{a})^2 f(a)da. \tag{30}$$

Observe that agents can be now split into two categories. Those who remain at their original choices when $\tau_\lambda(\,)$ is replaced with $\tau_{\bar{y}_R}(\,)$, and those who alter their behavior. Specifically, when the behavior is not changed, then by definition, we have $y_a^{**} = y_a^*$, and consequently it must be

$$\frac{1}{2}(\frac{a^2 - y_a^{**}}{a})^2 = \frac{1}{2}(\frac{a^2 - y_a^*}{a})^2. \tag{31}$$

It is necessary to consider two separate cases when the behavior is affected. First note that agents who initially chose incomes below $\bar{y}_R$ and now choose incomes above $\bar{y}_R$ pay more in taxes, i.e., we must have

$$R_a^{**}(\lambda, \bar{y}_R) = \tau_{\bar{y}_R}(y_a^{**}) > \tau_\lambda(y_a^*) = R_a^* + \lambda. \tag{32}$$

Furthermore, their previous choices remain available, so invoking the revealed preference argument we can write

$$h(y_a^{**} - R_a^{**}(\lambda, \bar{y}_R) - \frac{1}{2}(\frac{y_a^{**}}{a})^2) \geq h(y_a^* - R_a^* - \lambda - \frac{1}{2}(\frac{y_a^*}{a})^2), \tag{33}$$

which can be simplified to

$$y_a^{**} - R_a^{**}(\lambda, \bar{y}_R) - \frac{1}{2}(\frac{y_a^{**}}{a})^2 \geq y_a^* - R_a^* - \lambda - \frac{1}{2}(\frac{y_a^*}{a})^2. \tag{34}$$

In turn, by adding up (32) and (34), we can establish the following

$$y_a^{**} - \frac{1}{2}(\frac{y_a^{**}}{a})^2 > y_a^* - \frac{1}{2}(\frac{y_a^*}{a})^2, \tag{35}$$

which naturally implies that

$$\frac{1}{2}(\frac{a^2 - y_a^*}{a})^2 > \frac{1}{2}(\frac{a^2 - y_a^{**}}{a})^2. \tag{36}$$

The situation is slightly more complicated in the case of agents who initially chose incomes above $\bar{y}_R$. First of all, note that such agents would never switch to levels of income below $\bar{y}_R$, as those choices were available to them originally, and they chose incomes above $\bar{y}_R$. Moreover, now, they can actually pay less in taxes if they remain above $\bar{y}_R$, so switching to income levels below $\bar{y}_R$ is surely suboptimal. Furthermore, remaining above $\bar{y}_R$ entails, first of all, a lower amount of tax liabilities, and secondly, it implies that agents face a marginal tax of zero. Consequently, we can conclude that agents who were initially above $\bar{y}_R$ remain in the range of incomes above $\bar{y}_R$ and, in fact, do change their behavior in response to zero marginal taxes. In other words, the new labor supply choices coincide with the values of labor supply that would be chosen if taxes were lump sum. Accordingly, we must have

$$h(y_a^{**} - R_a^* - \lambda - \frac{1}{2}(\frac{y_a^{**}}{a})^2) > h(y_a^* - R_a^* - \lambda - \frac{1}{2}(\frac{y_a^*}{a})^2), \tag{37}$$

as supplying a given level of revenue, $R_a^* + \lambda$, is more efficient when marginal taxes are zero rather than positive. Equation (37) naturally reduces to

$$y_a^{**} - \frac{1}{2}(\frac{y_a^{**}}{a})^2 > y_a^* - \frac{1}{2}(\frac{y_a^*}{a})^2 \tag{38}$$

and further to

$$\frac{1}{2}(\frac{a^2 - y_a^*}{a})^2 > \frac{1}{2}(\frac{a^2 - y_a^{**}}{a})^2. \tag{39}$$

Combining (31), (36), and (38), we can be sure that the following must be true

$$\forall a \in [a_L, a_H] | \ \frac{1}{2}(\frac{a^2 - y_a^*}{a})^2 \geq \frac{1}{2}(\frac{a^2 - y_a^{**}}{a})^2, \tag{40}$$

with a strict inequality on a non-degenerate set. Furthermore, the collection of inequalities (40) implies that

$$\int_{a_L}^{a_H} \frac{1}{2}(\frac{a^2 - y_a^*}{a})^2 f(a)da > \int_{a_L}^{a_H} \frac{1}{2}(\frac{a^2 - y_a^{**}}{a})^2 f(a)da. \tag{41}$$

Observe that by replacing the efficient tax function $\tau_E(\,)$ with $\tau_\lambda(\,) = \tau_E(\,) + \lambda$ for points actually chosen, we increase the revenue collected, and we do not affect individual choices, i.e., we do not affect efficiency. Furthermore, by replacing $\tau_\lambda(\,) = \tau_E(\,) + \lambda$ with $\tau_{\bar{y}_R}(\,)$, we enhance efficiency and reduce the revenue collected to the level obtained with $\tau_E(\,)$. Clearly, as we replace $\tau_E(\,)$ with $\tau_{\bar{y}_R}(\,)$ we improve efficiency without compromising revenue. In other words, it is possible to collect, in more efficient way, the same amount of revenue as collected with $\tau_E(\,)$.

We can restate the argument formally, as we can express the original, when agents face $\tau_E(\,)$, value of the dead weight burden, equation (18), as

$$DWB^* = \int_{a_L}^{a_H} \frac{1}{2}(\frac{a^2 - y_a^*}{a})^2 f(a)da. \tag{42}$$

Similarly, the dead weight burden when agents face $\tau_{\bar{y}_R}(\,)$ is given by, equation (30) – restated below,

$$DWB^{**} = \int_{a_L}^{a_H} \frac{1}{2}(\frac{a^2 - y_a^{**}}{a})^2 f(a)da. \tag{43}$$

However, given (41), we can easily establish that

$$DWB^{**} < DWB^*. \tag{44}$$

Therefore, we must conclude that the original function, $\tau_E(\,)$, cannot be efficient, as by replacing $\tau_E(\,)$ with $\tau_{\bar{y}_R}(\,)$, we not only preserve revenue, but also enhance efficiency since we lower the value of the objective functional expressed with (14). In other words, a given tax function cannot be efficient unless it is flat in a neighborhood of the highest income earned. Consequently, we can state that a given tax function cannot be efficient unless it is of the shape of $\tau_{\bar{y}_R}(\,)$ to start with. However, by originating at y_H and moving to the left, we can repeat the above argument starting from any point y that is actually chosen and at which the efficient tax function stops being flat. Consequently, we can argue[6] that any efficient tax function must be a step function at least over an non-degenerate set of income levels.

To reiterate we can state that it turns out that the class of functions implicitly considered plausible by Saez [7] can be too limited. Specifically, we can apply Saez's approach even when the distribution of skills is bounded. Let $\tau_S()$ be the efficient tax function obtained with Saez's method. Naturally, we can always apply our procedure[7] to $\tau_S()$. First we shift it up by λ, small enough, and then we flatten its right hand tail to ensure revenue neutrality. Naturally, by doing so, we obviously enhance efficiency, i.e., we make the value of the total dead weight burden smaller. In other words, it is possible to improve[8] upon Saez's solution if one is willing to extend the class of

[6]We provide precise reasoning in Appendix C.

[7]Appendix E considers the case when some agents choose not to work when they face $\tau_S(\,)$.

[8]Note that we initially assumed that the revenue requirement, R, was sufficiently high to exclude

admissible tax functions and, in particular, consider step functions as admissible.

Furthermore, Saez [7] argues that his dead weight burden minimization problem is equivalent to an optimal taxation problem with properly defined social welfare weights. However, this signals a possibility that solutions to some of the optimal taxation problems can be improved upon by extending the class of admissible functions.

4 Preferences for Redistribution

In his original contribution Mirrlees [4] noted that policy makers apart from being interested in efficiency issues can also take into account equity considerations. In this paper, we adopt Mirrlees's approach, but we do not assume that the preferences of policy makers are defined over utility levels, but rather over the degree of inefficiency created by tax policies.

As previously let us assume that given tax function $\tau(\,)$ the problem of a given economic agent can be stated as

$$max_{\{y\}}U = h(y - \tau(y) - \frac{1}{2}(\frac{y}{a})^2). \tag{45}$$

Denoting with y_a the optimal choice we can express the value of the dead weight burden as

$$DWB_a = \frac{1}{2}(\frac{a^2 - y_a}{a})^2. \tag{46}$$

We can now define the problem of a policy maker as follows

$$min_{\{\tau(\,)\}}W = \int_{a_L}^{a_H} G(DWB_a)f(a)da \tag{47}$$

subject to

$$R = \int_{a_L}^{a_H} \tau(y_a)f(a)da \tag{48}$$

and

$$\forall a \in [a_L, a_H] |\, h^{'}(y_a - \tau(y_a) - \frac{1}{2}(\frac{y_a}{a})^2)(1 - \tau'(y_a) - \frac{y_a}{a^2}) = 0. \tag{49}$$

Naturally, function $G(\,)$ captures the preferences of the policy maker. Following

---

the feasibility of a trivial solution in the form of a uniform tax of $\frac{R}{\int_{a_L}^{a_H} f(a)da}$. Consequently, our improvement is applicable in cases, considered of interest by Saez [7], when some agents are not able to pay a uniform poll tax.

Mirrlees we assume that $G(\,)$ is increasing and convex[9], even though for our purposes only the former condition is relevant. Observe that our formulation of the problem is different than that proposed by Mirrlees. Specifically, in our case $G()$ is assumed to be a function of the dead weight burden whereas Mirrlees assumes that policy makers are utilitarian. In other words in our case policy makers are assumed to exhibit aversion towards variations in the degree of inefficiency faced by economic agents rather than in the *level* of overall happiness.

We propose this alternative formulation for a number of reasons. First of all, it appears that the literature has so far ignored the problem, which appears to be of relevance. Secondly, the problem as proposed by Mirrlees [4] appears to have been tackled given the advances of Diamond [1] and Saez [6]. Finally, our formulation, we believe, in addition to its novelty is more natural than that proposed by Mirrlees. Note that in Mirrlees's approach one must accept the notion that utility levels are meaningful and more importantly that cross person comparison of utilities can be performed. Those implicit assumptions of Mirrlees are normally taken as given and are normally not questioned. However, formally they contradict some of the basic foundations of economic theory. In our formulation, on the other hand, those standard problems do not arise as dead weight burdens for all individuals are measured in common units, which makes ranking of outcomes and cross personal comparisons meaningful. In other words, the problem as stated in not only novel, but it is also well defined and it does not contradict any of the primitive assumptions of economic theory.

Let $\tau_G(\,)$ be the solution, assuming that it exists, to the problem described with equations (47), (48), and (49). Furthermore, let y_a^* denote that the optimal choice of an agent whose productivity is equal to a when she faces tax function $\tau_G(\,)$.

At this stage we cannot say anything meaningful about the shape of $\tau_G(\,)$ apart from some basic observations[10]. However, let us assume that $\tau_G()$ actually exists and, for illustrative purposes, that its shape resembles the shape of optimal tax functions. Furthermore, recall that in the main body of the paper we focus on the case when all

[9] Note that we are interested in minimization of the dead weight burden. Consequently, working a convex rather than concave function $G(\,)$ appears to be more proper if one intends to capture aversion towards inequality.

[10] It is straightforward to establish that in this case just as it is in the case of optimal taxation the marginal tax paid by the most talented individual should be zero.

agents, given $\tau_G()$, strictly prefer to supply a positive amount of labor[11]. Under those assumptions the dead weight burden experienced by an agent whose productivity is equal to a can be expressed as

$$DWB_a^* = \frac{1}{2}(\frac{a^2 - y_a^*}{a})^2 \tag{50}$$

and the corresponding value of the objective functional is given by

$$W^* = \int_{a_L}^{a_H} G(\frac{1}{2}(\frac{a^2 - y_a^*}{a})^2) f(a) da. \tag{51}$$

Let us now consider familiar variations of $\tau_G(\,)$. First let us shift $\tau_G(\,)$ over the relevant range up by λ, i.e., let us consider $\tau_\lambda(y) = \tau_G(y) + \lambda$ for values of y that are actually chosen[12]. Note that such a transformation has no bearing on individual behavior for λ sufficiently small. Furthermore, let us replace $\tau_\lambda()$ with a tax function, $\tau_{\bar{y}}(\,)$, that coincides with $\tau_\lambda(\,)$ up to a given level of income, $\bar{y}$, and then it is flat afterwards. Recall that $\tau_\lambda(\,)$ brings more revenue than $\tau_G(\,)$ and that we can find such a value, $\bar{y}_R$, see the discussion in Appendix B, that the level of revenue collected with $\tau_{\bar{y}_R}(\,)$ is the same as the level of revenue collected with $\tau_G(\,)$. Formally, we can state that there exists $\bar{y}_R$ such that $\tau_{\bar{y}_R}(\,)$ is feasible for the problem described with (47), (48), and (49), which can be stated as

$$\int_{a_L}^{a_H} \tau_G(y_a^*) f(a) da = \int_{a_L}^{a_H} \tau_{\bar{y}_R}(y_a^{**}) f(a) da = R, \tag{52}$$

where y_a^{**} denotes the level of income earned by an agent whose productivity is equal to a when she faces tax function $\tau_{\bar{y}_R}(\,)$.

As argued in the previous section the dead weight burden experienced by an individual when she faces $\tau_{\bar{y}_R}(\,)$ is never higher than the level of dead weight burden when she faces $\tau_G(\,)$. Consequently, we can write that

$$\forall a \in [a_L, a_H] | \; \frac{1}{2}(\frac{a^2 - y_a^*}{a})^2 \geq \frac{1}{2}(\frac{a^2 - y_a^{**}}{a})^2, \tag{53}$$

with a strict in equality on a non-degenerate set.

[11] Again, we consider the case when some agents are idle in Appendix E.

[12] We can assume that $\tau_\lambda(y) = y$ everywhere else.

Furthermore, note that by assumption $G(\,)$ is an increasing function. Consequently, given series of inequalities (53) we can write the following

$$\forall a \in [a_L, a_H]|\; G(\frac{1}{2}(\frac{a^2 - y_a^*}{a})^2) \geq G(\frac{1}{2}(\frac{a^2 - y_a^{**}}{a})^2), \tag{54}$$

which implies, as the inequality in (54) strict for a non-degenerate set of a, that the following holds

$$W^* = \int_{a_L}^{a_H} G(\frac{1}{2}(\frac{a^2 - y_a^*}{a})^2) f(a) da > \int_{a_L}^{a_H} G(\frac{1}{2}(\frac{a^2 - y_a^{**}}{a})^2) f(a) da = W^{**}. \tag{55}$$

Naturally, inequality (55) together with the feasibility condition (52) imply that $\tau_G(\,)$ cannot be a solution to the problem described with (47), (48), and (49) unless it is locally flat in a neighborhood of the highest income earned for otherwise it could be possible to lower the value of the objective functional by flattening $\tau_G(\,)$ at the high end.

The solution to the problem described with (47), (48), and (49) is again particularly simple. It turns out that allowing the policy maker to display aversion towards inequality does not change the solution in the qualitative sense. It is still the case that the solution is of the form of a step function. In other words, the preferences of the policy maker do not affect the form of the solution, but naturally determine its precise shape.

5 Conclusions

In this paper, we focus on the efficiency of taxation. Specifically, we modify the optimal taxation problem of Mirrlees [4] by assuming that the preferences of policy makers are defined over the magnitude of distortions created by tax systems rather than the overall level of utility. In particular, we choose to measure distortions using the most natural metric, i.e., the dead weight burden imposed on individual agents by tax systems. Furthermore, we allow policy makers to exhibit aversion towards inequality, i.e., we assume that policy makers value outcomes that are characterized by a lesser dispersion of distortions.

We consider our approach and our results to be complementary to those studied and derived in the standard optimal taxation literature. At the technical level our modification amounts to a change in the form of the objective functional. The change that we propose is a very natural one and it allows us to extend the optimal taxation literature in a meaningful manner. Specifically, by assuming that the preferences of policy makers are defined over the magnitudes of dead weight burden rather than over the levels of utility we eliminate the most profound criticism formulated with regard to the optimal taxation approach. Our specification and focus on dead weight burden rather than the level of utility makes comparison of utilities meaningful and allows for ranking of the values of the objective functional in a manner that is deeply rooted and consistent with basic economic theory. Furthermore, in our case the preferences of the policy maker are defined over concepts measured in common and comparable units making comparisons across individuals informative. Consequently, the extension that we propose is not only novel, but, in fact, it can be justified by fundamental assumptions of economic theory and not just by the applicability of a given method that facilitates the solution of the problem at hand.

In the paper, at the technical level, we work, following Diamond [1] and Saez [6], with a utility function that is essentially of a quasi-linear form, and we show, contrary to the findings of Saez [7], that tax functions that ensure efficiency are of particularly simple form. Specifically, we show that efficient tax functions, if they exist, are locally flat implying generically zero marginal tax rates. In addition, we prove that our findings hold even when policy makers exhibit aversion towards inequality.

We view our contribution to be of value at a number of levels. First of all, we propose a framework that is, unlike the optimal taxation literature, fully compatible with basic economic theory. Secondly, while the approach of Mirrlees [4] has become dominant it is not the only one. Recently, Werning [12], and Saez and Stantcheva [8] have noted the limitations of the approach of Mirrlees and have proposed alternative formulations of the optimal taxation problems. In this paper, we follow a similar path; we propose an alternative, which is very natural, and leads to new and meaningful results.

References

[1] Diamond, Peter A., (1998). "Optimal Income Taxation: An Example with a U-Shaped pattern of Optimal analysis Tax Rates," American Economic Review, 88, pp. 83-95.

[2] Ebert, Udo, (1992). "A Reexamination of the Optimal Nonlinear Income Tax," Journal of Public Economics, 49, pp. 47-73.

[3] Lollivier, Stefan and Jean-Charles Rochet, (1983). "Bunching and Second Order Conditions: a Note on Optimal Tax Theory," Journal of Economic Theory, 31, pp. 392-400.

[4] Mirrlees, James A., (1971). "An Exploration in the Theory of Optimal Income Taxation," Review of Economic Studies, 38, pp. 175-208.

[5] Sadka, Efraim, (1976). "On Income Distribution, Incentive Effects and Optimal Income Taxation," Review of Economic Studies, 42, pp. 261-268.

[6] Saez, Emmanuel, (2001). "Using Elasticities to Derive Optimal Income tax Rates," Review of Economic Studies, 68, pp. 205-229.

[7] Saez, Emmanuel, (1999). "A Characterization of the Income Tax Schedule Minimizing Deadweight Burden," Chapter 3, MIT Ph.D. Thesis.

[8] Saez, Emmanuel, and Stafanie Stantcheva, (2012). "Optimal Tax Theory with Endogenous Social Marginal Welfare Weights," UC Berkeley, mimeo.

[9] Seade, Jesus K., (1977). "On the Shape of Optimal Tax Schedules," Journal of Public Economics, 7, pp. 203-236.

[10] Seade, Jesus K., (1982). "On the Sign of the Optimum Marginal Income Tax," Review of Economic Studies, 49, pp. 637-643.

[11] Stiglitz, Joseph E., (1987). "Pareto Efficient and Optimal Taxation and the New New Welfare Economics," Handbook of Public Economics, Volume 2, edited by A.J. Auerbach and M. Feldstein, Elsevier Science Publishers B.V. (North Holland)

[12] Werning, Ivan, (2007). "Pareto Efficient Income Taxation," MIT, mimeo.

A Appendix

It can be the case that some values of income are not chosen when economic agents face a given efficient tax function, $\tau()$. In this appendix, we describe our basic variation

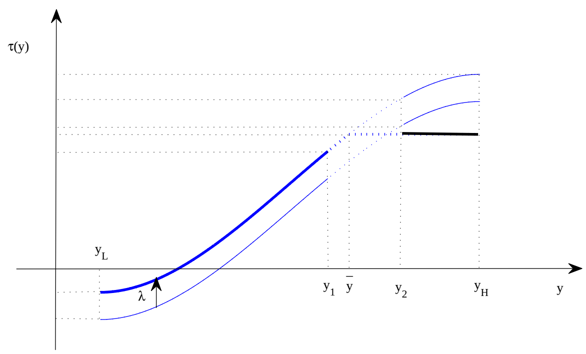

Figure 4: The form of $\tau_{\bar{y}_R}(\ )$ when some values of income are not chosen.

of the optimal tax function in the case when there are unchosen ranges of income. Specifically, let us assume that the range, (y_1, y_2), is the highest range of income that is unchosen, which we illustrate in figure (4) with a broken line.

Again, let us assume that we first shift the efficient tax function, $\tau_E(\ )$, upward by λ, i.e., we impose an additional lump sum tax on all agents of λ. Secondly, we start flattening the far end of the optimal tax function starting from $\bar{y}$ in search of $\bar{y}_R$, which would ensure revenue neutrality as compared to the amount of revenue collected with $\tau_E(\ )$. If it turns out that $\bar{y}$ reaches y_2 before we attain neutrality, we continue flattening of the tax function, but we only allow for income levels that were chosen initially; i.e., we do not allow agents to choose incomes from the range that was not initially chosen, (y_1, y_2). We depict the process in figure (4). Furthermore, if revenue neutrality is not obtained when $\bar{y}$ reaches y_1 then we continue the process further by flattening the part of the tax function starting from y_1 onwards.

In the next appendix we show that a variation of the form described above leads to a continuous change in the amount of revenue collected. Therefore, we can expect that there is a value of $\bar{y}$ at which the revenue collected is equal to that originally raised with the efficient tax function, $\tau_E(\ )$.

B Appendix

It is our goal, in this Appendix, to establish that the variational procedures considered in this paper lead to a continuous change in the amount of revenue collected by the government. We start from the case when all values of income are chosen in a neighborhood of the highest income earned y_H. Moreover, let us consider a tax function, $\tau(\ )$, of the form given in figure (1) and let us assume that

$$\forall y \in [y_L, y_H]|\ |\tau(y)| < T \tag{56}$$

and

$$\forall a \in [a_L, a_H]|\ |f(a)| < \bar{f}. \tag{57}$$

Let us consider range of income levels $[y_M, y_H]$ and let us assume that $\tau(\ )$ is concave on $[y_M, y_H]$. Moreover, let us assume that all income levels from $[y_M, y_H]$ are actually chosen. Let us consider an agent whose productivity is equal to a who chooses $y_a^* \in [y_M, y_H]$. Naturally, efficiency requires that

$$1 - \tau'(y_a^*) = \frac{y_a^*}{a^2} \tag{58}$$

and

$$-\tau''(y_a^*) - \frac{1}{a^2} < 0. \tag{59}$$

Naturally, inequality (59) together with the assumption of concavity of $\tau()$ on $[y_M, y_H]$ imply that

$$-\frac{1}{a^2} < \tau''(y_a^*) < 0. \tag{60}$$

Consequently, the series of inequalities (60) allows us to establish that

$$\forall a|\ y_a^* \in [y_M, y_H] \Rightarrow |\tau''(y_a^*)| < \frac{1}{a^2}. \tag{61}$$

Now given our assumption that all incomes from $[y_M, y_H]$ are actually chosen we can rewrite (61) as

$$\forall y \in [y_M, y_H]|\ |\tau''(y)| < \frac{1}{a_L^2}, \tag{62}$$

i.e., the second derivative of $\tau(\ )$ remains bounded from above in absolute value.

Furthermore, given equation (58) we can write inequality (59) as

$$0 < \tau''(y_a^*) + \frac{1 - \tau'(y_a^*)}{y_a^*}. \tag{63}$$

Naturally, given that $\tau''(\,)$ and $\tau'(\,)$ exist the following function

$$m(y) = \tau''(y) + \frac{1 - \tau'(y)}{y} \tag{64}$$

is continuous. Therefore, we can write that

$$\forall y \in [y_M, y_H] |\; 0 < M_L \leq \tau''(y) + \frac{1 - \tau'(y)}{y} \leq M_H \tag{65}$$

since any continuous function reaches its bounds on a closed interval and inequality (63) holds for all $y_a^* \in [y_M, y_H]$.

Now, let us focus on the following modification of $\tau(\,)$

$$\tau_{\bar{y}}(y) = \begin{cases} \tau(y) + \lambda & \text{for } y \leq \bar{y} \\ \tau(\bar{y}) + \lambda & \text{for } \bar{y} < y \end{cases}, \tag{66}$$

where $\bar{y}$ is assumed to be close enough to y_H.

Furthermore, let y_a^* denote the income level chosen by an agent whose productivity is equal to a when she faces $\tau(\,)$ and y_a^{**} be the level of income chosen by an agent whose productivity is equal to a when she faces $\tau_{\bar{y}}(\,)$. Naturally, in the latter case the revenue by the government collected is given by $R(\bar{y}) = \int_{a_L}^{a_H} \tau_{\bar{y}}(y_a^{**}) f(a) da$. Let $\tilde{a}$ denote the level of productivity at which agents are just indifferent between remaining on the steep part of $\tau_{\bar{y}}(\,)$ or on the flat part of $\tau_{\bar{y}}(\,)$. Naturally, we must have

$$h(y_{\tilde{a}}^* - \tau(y_{\tilde{a}}^*) - \lambda - \frac{1}{2}(\frac{y_{\tilde{a}}^*}{\tilde{a}})^2) = h(\frac{1}{2}\tilde{a}^2 - \tau(\bar{y}) - \lambda), \tag{67}$$

which obviously reduces to

$$y_{\tilde{a}}^* - \tau(y_{\tilde{a}}^*) - \frac{1}{2}(\frac{y_{\tilde{a}}^*}{\tilde{a}})^2 = \frac{1}{2}\tilde{a}^2 - \tau(\bar{y}). \tag{68}$$

Naturally, $y^*_{\tilde{a}}$ is an optimal choice, so it must be the case that

$$1-\tau'(y^*_{\tilde{a}})=\frac{y^*_{\tilde{a}}}{\tilde{a}^2} \tag{69}$$

and

$$-\tau''(y^*_{\tilde{a}})-\frac{1}{\tilde{a}^2}<0. \tag{70}$$

Observe that we can express the revenue collected in this case as

$$R(\bar{y})=\int_{a_L}^{\tilde{a}}\tau(y^*_a)f(a)da+\tau(\bar{y})\int_{\tilde{a}}^{a_H}f(a)da+\lambda\int_{a_L}^{a_H}f(a)da. \tag{71}$$

Let us now consider a simple variation of $\tau_{\bar{y}}(\,)$ of the form

$$\tau_{\bar{y}-\varepsilon}(y)=\begin{cases}\tau(y)+\lambda & \text{for } y\leq\bar{y}-\varepsilon\\ \tau(\bar{y}-\varepsilon)+\lambda & \text{for } \bar{y}-\varepsilon<y\end{cases}. \tag{72}$$

Naturally, $\tau_{\bar{y}-\varepsilon}(\,)$ is just the same as $\tau(\,)+\lambda$ with its end flattened starting from $\bar{y}-\varepsilon$.

Let $\tilde{a}_\varepsilon$ be the value of productivity at which agents are just indifferent between paying a positive marginal tax and zero marginal tax when they face $\tau_{\bar{y}-\varepsilon}()$. Obviously, we must have

$$y^*_{\tilde{a}_\varepsilon}-\tau(y^*_{\tilde{a}_\varepsilon})-\frac{1}{2}(\frac{y^*_{\tilde{a}_\varepsilon}}{\tilde{a}_\varepsilon})^2=\frac{1}{2}\tilde{a}_\varepsilon^2-\tau(\bar{y}-\varepsilon). \tag{73}$$

Furthermore, optimality of $y^*_{\tilde{a}_\varepsilon}$ requires that

$$1-\tau'(y^*_{\tilde{a}_\varepsilon})=\frac{y^*_{\tilde{a}_\varepsilon}}{\tilde{a}_\varepsilon^2}, \tag{74}$$

and

$$-\tau''(y^*_{\tilde{a}_\varepsilon})-\frac{1}{\tilde{a}_\varepsilon^2}<0. \tag{75}$$

The level of revenue collected in this case is given by

$$R(\bar{y}-\varepsilon)=\int_{a_L}^{\tilde{a}_\varepsilon}\tau(y^*_a)f(a)da+\tau(\bar{y}-\varepsilon)\int_{\tilde{a}_\varepsilon}^{a_H}f(a)da+\lambda\int_{a_L}^{a_H}f(a)da. \tag{76}$$

We can using (71) and (76) estimate the size of the difference between the two levels of revenue collected. In particular, we have

$$|R(\bar{y})-R(\bar{y}-\varepsilon)| = |\int_{\tilde{a}_\varepsilon}^{\tilde{a}} (\tau(y_a^*)-\tau(\bar{y}-\varepsilon))f(a)da+(\tau(\bar{y})-\tau(\bar{y}-\varepsilon))\int_{\tilde{a}}^{a_H} f(a)da|, \quad (77)$$

which can be bounded from above as follows

$$|R(\bar{y})-R(\bar{y}-\varepsilon)| \leq \int_{\tilde{a}_\varepsilon}^{\tilde{a}} |\tau(y_a^*)-\tau(\bar{y}-\varepsilon)|f(a)da+|\tau(\bar{y})-\tau(\bar{y}-\varepsilon)|\int_{\tilde{a}}^{a_H} f(a)da. \quad (78)$$

Furthermore, given that $\tau(\,)$ is bounded, restriction (56), and $f(\,)$ is bounded, restriction (57), we can write

$$|R(\bar{y})-R(\bar{y}-\varepsilon)| \leq |\tilde{a}-\tilde{a}_\varepsilon|2T\bar{f}+|\tau(\bar{y})-\tau(\bar{y}-\varepsilon)|\int_{\tilde{a}}^{a_H} f(a)da. \quad (79)$$

Now using equations (69) and (74) we can establish that

$$|\frac{1}{\tilde{a}_\varepsilon^2}-\frac{1}{\tilde{a}^2}| = |\frac{1-\tau'(y_{\tilde{a}_\varepsilon}^*)}{y_{\tilde{a}_\varepsilon}^*}-\frac{1-\tau'(y_{\tilde{a}}^*)}{y_{\tilde{a}}^*}| = |\frac{-\tau''(y_c)y_c-(1-\tau'(y_c))}{y_c^2}||y_{\tilde{a}_\varepsilon}^*-y_{\tilde{a}}^*|, \quad (80)$$

where $y_c \in [y_{\tilde{a}_\varepsilon}^*, y_{\tilde{a}}^*]$.

Now, invoking inequality (62) and relying on $0 \leq |\tau'(\,)| \leq 1|$, finding due to Seade [9, 10], and noting that $y_c \in [y_L, y_H]$ we can use equation (80) to establish the following

$$|\frac{1}{\tilde{a}_\varepsilon^2}-\frac{1}{\tilde{a}^2}| < \frac{\frac{1}{a_L^2}y_H+1}{y_L^2}|y_{\tilde{a}_\varepsilon}^*-y_{\tilde{a}}^*|, \quad (81)$$

which is equivalent to

$$|\tilde{a}_\varepsilon-\tilde{a}| < \frac{(\tilde{a}_\varepsilon\tilde{a})^2}{\tilde{a}_\varepsilon+\tilde{a}}\frac{\frac{1}{a_L^2}y_H+1}{y_L^2}|y_{\tilde{a}_\varepsilon}^*-y_{\tilde{a}}^*|, \quad (82)$$

which in turn implies that

$$|\tilde{a}_\varepsilon-\tilde{a}| < \frac{a_H^4}{2a_L}\frac{\frac{1}{a_L^2}y_H+1}{y_L^2}|y_{\tilde{a}_\varepsilon}^*-y_{\tilde{a}}^*|. \quad (83)$$

Observe that we can use equations (68) and (69) to establish that

$$y^*_{\tilde{a}} - \tau(y^*_{\tilde{a}}) - \frac{1}{2}y^*_{\tilde{a}}(1-\tau'(y^*_{\tilde{a}})) - \frac{1}{2}\frac{y^*_{\tilde{a}}}{1-\tau'(y^*_{\tilde{a}})} = -\tau(\bar{y}). \tag{84}$$

Similarly, using equations (73) and (74) we can show that

$$y^*_{\tilde{a}_\varepsilon} - \tau(y^*_{\tilde{a}_\varepsilon}) - \frac{1}{2}y^*_{\tilde{a}_\varepsilon}(1-\tau'(y^*_{\tilde{a}_\varepsilon})) - \frac{1}{2}\frac{y^*_{\tilde{a}_\varepsilon}}{1-\tau'(y^*_{\tilde{a}_\varepsilon})} = -\tau(\bar{y}-\varepsilon). \tag{85}$$

Now, by subtracting equation (84) from equation (85) and using Taylor expansion we can establish that

$$\frac{1}{2}|1-\tau'(y^*_D) + y^*_D\tau''(y^*_D)||1 - \frac{1}{(1-\tau'(y^*_D))^2}||y^*_{\tilde{a}_\varepsilon} - y^*_{\tilde{a}}| = |\tau(\bar{y}) - \tau(\bar{y}-\varepsilon)|, \tag{86}$$

where $y^*_D \in [y^*_{\tilde{a}_\varepsilon}, y^*_{\tilde{a}}]$.

We can rearrange equation (86) to

$$\frac{1}{2}|\frac{1-\tau'(y^*_D)}{y^*_D} + \tau''(y^*_D)||\frac{y^*_D}{1-\tau'(y^*_D)}||\tau^{'}(y^*_D)||1 + \frac{1}{1-\tau'(y^*_D)}||y^*_{\tilde{a}_\varepsilon} - y^*_{\tilde{a}}| = |\tau(\bar{y}) - \tau(\bar{y}-\varepsilon)|. \tag{87}$$

Now, equation (87) implies, given (65), that

$$\frac{1}{2}M_L|\frac{y^*_D}{1-\tau'(y^*_D)}||\tau^{'}(y^*_D)||1 + \frac{1}{1-\tau'(y^*_D)}||y^*_{\tilde{a}_\varepsilon} - y^*_{\tilde{a}}| \leq |\tau(\bar{y}) - \tau(\bar{y}-\varepsilon)|. \tag{88}$$

Recall that by assumption $\tau''(y)$ is negative for $y \in [y_M, y_H]$ therefore $\tau'(\ )$ is an decreasing function on $[y_M, y_H]$. Thus, we can write

$$1 > \tau'(y^*_D) > \tau'(\bar{y}) > 0 \tag{89}$$

since $y^*_D \in [y^*_{\tilde{a}_\varepsilon}, y^*_{\tilde{a}}]$, and $y^*_{\tilde{a}} < \bar{y}$, and optimal marginal taxes are smaller than 1, see Seade [9, 10].

Naturally (89) is equivalent to

$$0 < 1 - \tau'(y^*_D) < 1 - \tau'(\bar{y}) < 1, \tag{90}$$

which in turn implies that

$$1 < \frac{1}{1-\tau'(\bar{y})} < \frac{1}{1-\tau'(y_D^*)}. \tag{91}$$

Obviously, the above inequalities lead to

$$2 < |1 + \frac{1}{1-\tau'(y_D^*)}|. \tag{92}$$

Thus we can inequality (88) can be rewritten as

$$M_L|\frac{y_D^*}{1-\tau'(y_D^*)}||\tau^{'}(y_D^*)||y_{\tilde{a}_\varepsilon}^* - y_{\tilde{a}}^*| \leq |\tau(\bar{y}) - \tau(\bar{y}-\varepsilon)|. \tag{93}$$

Now, invoking inequalities (89) we can rewrite the above inequality as

$$M_L|\frac{y_D^*}{1-\tau'(y_D^*)}||\tau^{'}(\bar{y})||y_{\tilde{a}_\varepsilon}^* - y_{\tilde{a}}^*| \leq |\tau(\bar{y}) - \tau(\bar{y}-\varepsilon)|. \tag{94}$$

Furthermore, given our initial assumption, y_D^* is actually chosen by someone, i.e., we have

$$1-\tau'(y_D^*) = \frac{y_D^*}{a_D^2}. \tag{95}$$

Thus,

$$a_D^2 = \frac{y_D^*}{1-\tau'(y_D^*)} \tag{96}$$

and, hence, inequality (94) becomes

$$M_L a_D^2|\tau^{'}(\bar{y})||y_{\tilde{a}_\varepsilon}^* - y_{\tilde{a}}^*| \leq |\tau(\bar{y}) - \tau(\bar{y}-\varepsilon)|, \tag{97}$$

which in turn leads to, as $a_L < a_D$,

$$|y_{\tilde{a}_\varepsilon}^* - y_{\tilde{a}}^*| \leq \frac{|\tau(\bar{y}) - \tau(\bar{y}-\varepsilon)|}{M_L a_L^2|\tau'(\bar{y})|}. \tag{98}$$

Therefore, combining inequality (83) with inequality (98) we can write the following

$$|\tilde{a}_\varepsilon - \tilde{a}| < \frac{a_H^4}{2a_L}\frac{\frac{1}{a_L^2}y_H + 1}{y_L^2}|y_{\tilde{a}_\varepsilon}^* - y_{\tilde{a}}^*| \leq \frac{a_H^4}{2a_L}\frac{\frac{1}{a_L^2}y_H + 1}{y_L^2}\frac{|\tau(\bar{y}) - \tau(\bar{y}-\varepsilon)|}{M_L a_L^2|\tau'(\bar{y})|}. \tag{99}$$

Noting that $y_H = a_H^2$, $y_L = a_L^2$, and $1 < \frac{a_H^2}{a_L^2}$ we can rewrite the above inequality as

$$|\tilde{a}_\varepsilon - \tilde{a}| \leq \frac{a_H^6}{a_L^7} \frac{|\tau(\bar{y}) - \tau(\bar{y} - \varepsilon)|}{M_L |\tau'(\bar{y})|}. \tag{100}$$

Furthermore, we can now rewrite inequality (79) as

$$|R(\bar{y}) - R(\bar{y} - \varepsilon)| \leq \frac{a_H^6}{a_L^7} \frac{|\tau(\bar{y}) - \tau(\bar{y} - \varepsilon)|}{M_L |\tau'(\bar{y})|} 2T\bar{f} + |\tau(\bar{y}) - \tau(\bar{y} - \varepsilon)| \int_{\tilde{a}}^{a_H} f(a)da, \tag{101}$$

which reduces, as $a_L \leq \tilde{a}$ and consequently $\int_{\tilde{a}}^{a_H} f(a)da \leq \int_{a_L}^{a_H} f(a)da$, to

$$|R(\bar{y}) - R(\bar{y} - \varepsilon)| \leq (\frac{a_H^6}{a_L^7} \frac{2T\bar{f}}{M_L |\tau'(\bar{y})|} + \int_{a_L}^{a_H} f(a)da) |\tau(\bar{y}) - \tau(\bar{y} - \varepsilon)| \tag{102}$$

Obviously, using Taylor expansion inequality (102) can be rewritten as

$$|R(\bar{y}) - R(\bar{y} - \varepsilon)| \leq (\frac{a_H^6}{a_L^7} \frac{2T\bar{f}}{M_L |\tau'(\bar{y})|} + \int_{a_L}^{a_H} f(a)da) |\tau'(\bar{y} - \theta_B \varepsilon)| \varepsilon, \tag{103}$$

where $\theta_B \in [0, 1]$.

Finally, following Seade [9, 10] and assuming that marginal tax rates are bounded from above and below we can conclude that $|\tau'(\bar{y} - \theta_B \varepsilon)| < 1$ and hence inequality (103) becomes

$$|R(\bar{y}) - R(\bar{y} - \varepsilon)| \leq \varepsilon (\frac{a_H^6}{a_L^7} \frac{2T\bar{f}}{M_L |\tau'(\bar{y})|} + \int_{a_L}^{a_H} f(a)da), \tag{104}$$

which confirms that $R(\bar{y})$ is a continuous function of $\bar{y}$ as the expression multiplying ε in (104) is a constant. Therefore, we can conclude that our variational procedure leads to a continuous change in the amount of revenue collected by the government if all values of income are chosen in the range, $[y_M, y_L]$.

We still need to consider the remaining case when some income levels are originally unchosen. Let (y_1, y_2) be the highest[13] range of unchosen incomes. Naturally, we can expect that bunching occurs in this case at y_1. Furthermore, let us consider

[13] We discuss the case when such a range does not exist in the next Appendix.

modifications of the original tax functions of the form, discussed in Appendix A,

$$\tau_{\bar{y}}(y) = \begin{cases} \tau(y) + \lambda & \text{for } y \leq y_1 \\ y & \text{for } y_1 < y \leq y_2 \\ \tau(\bar{y}) + \lambda & \text{for } y_2 < y \end{cases}, \tag{105}$$

where $\bar{y} \in (y_1, y_2]$.

Let us denotes the range of individual productivities that lead to bunching at y_1 with $[a_1, a_2]$. Naturally, we must, in particular, have

$$h(y_1 - \tau(y_1) - \lambda - \frac{1}{2}(\frac{y_1}{a_2})^2) = h(y_2 - \tau(\bar{y}) - \lambda - \frac{1}{2}(\frac{y_2}{a_2})^2), \tag{106}$$

which reduces to

$$y_1 - \tau(y_1) - \frac{1}{2}(\frac{y_1}{a_2})^2 = y_2 - \tau(\bar{y}) - \frac{1}{2}(\frac{y_2}{a_2})^2 \tag{107}$$

and, further, implies that

$$\frac{1}{a_2^2} = 2\frac{y_2 - y_1 + \tau(y_1) - \tau(\bar{y})}{y_2^2 - y_1^2}. \tag{108}$$

The amount of revenue collected when agents face $\tau_{\bar{y}}(\ )$ described with (105) is given by

$$R(\bar{y}) = \int_{a_L}^{a_1} \tau(y_a^*) f(a) da + \tau(y_1) \int_{a_1}^{a_2} f(a) da + \tau(\bar{y}) \int_{a_2}^{a_H} f(a) da + \lambda \int_{a_L}^{a_H} f(a) da. \tag{109}$$

Let us now consider the following modification of $\tau(y)$

$$\tau_{\bar{y}-\varepsilon}(y) = \begin{cases} \tau(y) + \lambda & \text{for } y \leq y_1 \\ y & \text{for } y_1 < y \leq y_2 \\ \tau(\bar{y} - \varepsilon) + \lambda & \text{for } y_2 < y \end{cases}. \tag{110}$$

Naturally, tax liabilities dictated by $\tau_{\bar{y}-\varepsilon}()$ above y_2 are smaller than those dictated by $\tau_{\bar{y}}(\ )$. Therefore, the range of productivities for which agents choose to bunch at y_1 shrinks. Let us denote the new range with $[a_1, a_2^{\varepsilon}] \subset [a_1, a_2]$.

Observe that in this case indifference between choosing $y_1$and y_2 by agents whose productivity is equal to a_2^ε implies that, an analog of (108),

$$\frac{1}{(a_2^\varepsilon)^2} = 2\frac{y_2 - y_1 + \tau(y_1) - \tau(\bar{y} - \varepsilon)}{y_2^2 - y_1^2}. \tag{111}$$

The level of revenue collected by the government when agents face $\tau_{\lambda,\bar{y}-\varepsilon}()$ is given by

$$R(\bar{y}) = \int_{a_L}^{a_1} \tau(y_a^*) f(a)da + \tau(y_1) \int_{a_1}^{a_2^\varepsilon} f(a)da + \tau(\bar{y} - \varepsilon) \int_{a_2^\varepsilon}^{a_H} f(a)da + \lambda \int_{a_L}^{a_H} f(a)da. \tag{112}$$

Now, using (109) and (112) we can establish that

$$|R(\bar{y}) - R(\bar{y}-\varepsilon)| = |(\tau(y_1) - \tau(\bar{y}-\varepsilon)) \int_{a_2^\varepsilon}^{a_2} f(a)da + (\tau(\bar{y}) - \tau(\bar{y}-\varepsilon)) \int_{a_2}^{a_H} f(a)da|, \tag{113}$$

which implies

$$|R(\bar{y}) - R(\bar{y}-\varepsilon)| \leq |\tau(y_1) - \tau(\bar{y}-\varepsilon)| \int_{a_2^\varepsilon}^{a_2} f(a)da + |\tau(\bar{y}) - \tau(\bar{y}-\varepsilon)| \int_{a_2}^{a_H} f(a)da \tag{114}$$

and in turn

$$|R(\bar{y}) - R(\bar{y} - \varepsilon)| \leq |a_2 - a_2^\varepsilon| 2T\bar{f} + |\tau^{'}(\bar{y} - \theta_C \varepsilon)| \varepsilon \int_{a_2}^{a_H} f(a)da, \tag{115}$$

where $\theta_c \in (0, 1)$, which further leads to, as marginal tax rates are bounded from above and below and $a_L < a_2$,

$$|R(\bar{y}) - R(\bar{y} - \varepsilon)| \leq |a_2 - a_2^\varepsilon| 2T\bar{f} + \varepsilon \int_{a_L}^{a_H} f(a)da. \tag{116}$$

Equations (108) and (111) imply that

$$\frac{1}{(a_2^\varepsilon)^2} - \frac{1}{a_2^2} = 2\frac{\tau(\bar{y}) - \tau(\bar{y} - \varepsilon)}{y_2^2 - y_1^2}, \tag{117}$$

which leads to

$$|a_2 - a_2^\varepsilon| = 2\frac{(a_2 a_2^\varepsilon)^2}{a_2 + a_2^\varepsilon} \frac{|\tau'(\bar{y} - \theta_D \varepsilon)| \varepsilon}{y_2^2 - y_1^2} \tag{118}$$

and, in turn, implies that

$$|a_2 - a_2^\varepsilon| \leq \frac{a_H^4}{a_L(y_2^2 - y_1^2)}\varepsilon. \tag{119}$$

Combining (116) and (119) we can establish that

$$|R(\bar{y}) - R(\bar{y} - \varepsilon)| \leq \varepsilon(\frac{a_H^4}{a_L(y_2^2 - y_1^2)}2T\bar{f} + \int_{a_L}^{a_H} f(a)da), \tag{120}$$

which confirms continuity of $R(\bar{y})$ since $\frac{a_H^4}{a_L(y_2^2 - y_1^2)}2T\bar{f} + \int_{a_L}^{a_H} f(a)da$ is a constant.

Therefore, we have established that the simple variational procedures applied in this text lead to a continuous change in the amount of revenue collected. Consequently, we can invoke mean values theorems and argue that

$$\exists \bar{y}_R \in [y_L, y_H]|\ R(\bar{y}_R) = R. \tag{121}$$

C Appendix

The basic procedure of flattening the far right hand end of the tax function, described in the main body of the text, can be normally extended to the entire domain. Before, we move with the proof let us make a technical observation. In principle, it could the case that the optimal tax function is characterized by multiple, possibly infinitely many, discontinuities. Furthermore, it could be the case that in any neighborhood of the highest income earned there are infinitely many discontinuities as illustrated in figure (5).

Naturally, if continuities are present and the function is locally flat in a neighborhood of a given discontinuity then there is nothing to prove as the optimal tax function is locally flat by assumption. On the other hand, if at some point the optimal tax function is no longer locally flat we can apply the procedure of flattening as described below. Specifically, the procedure that allows us to establish that the optimal tax function is locally flat consists in a simple repetition of the arguments presented above applied to the far right hand end of the optimal tax function. We show our basic point by invoking a proof by contradiction. Specifically, let us assume that the optimal tax function is not of the form of a step function. We know already that the optimal tax function must be flat in a neighborhood of the highest income

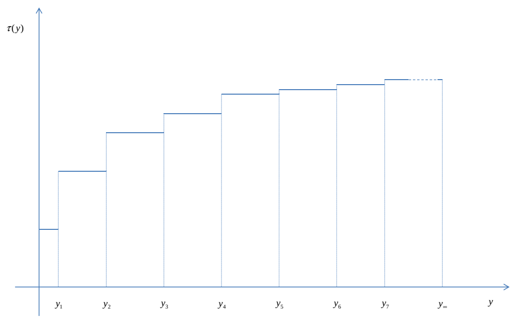

Figure 5: A Tax Function with Numerous Discontinuities.

earned. Consequently, we can expect it to take the form presented in figure (6).

Moreover, let us assume that y_1 is the highest value of income at which the efficient tax function starts being a step function or more precisely let y_1 be the highest value of income such that $\lim_{y \to y_1^+} \tau_E^{'}(y) \neq 0$. Furthermore, let $y_a^* = aL_a^*$ and $c_a^* = aL_a^* - \tau_E(aL_a^*)$ be the corresponding choices of an agent whose productivity is equal to a and who faces efficient tax function $\tau_E(\,)$. Naturally, the realized utility is in this case given by

$$U_a^* = h(y_a^* - \tau_E(y_a^*) - \frac{1}{2}(\frac{y_a^*}{a})^2). \tag{122}$$

Let us now consider a simple variation. Specifically, let us shift the part of the tax function to the left of y_1 up by λ, where λ is assumed to be sufficiently small, and let us leave the position of the remainder of the tax function unaffected. Formally, we define a new tax function as

$$\tau_\lambda(y) = \begin{cases} \tau_E(y) + \lambda & \text{for } y \leq y_1 \\ \tau_E(y) & \text{for } y > y_1 \end{cases}. \tag{123}$$

Observe that such a change in the shape of the tax function can induce some behavioral responses. Let us denote the new labor supply choices with L_a^{**}. Furthermore, observe that agents who initially chose income levels above y_1 do not change their behavior. Naturally, switching to values below y_1 cannot be optimal, as those choices were available up front, and now would only entail a higher level of tax liabilities. Furthermore, the form of the tax function above y_1 remains unchanged above

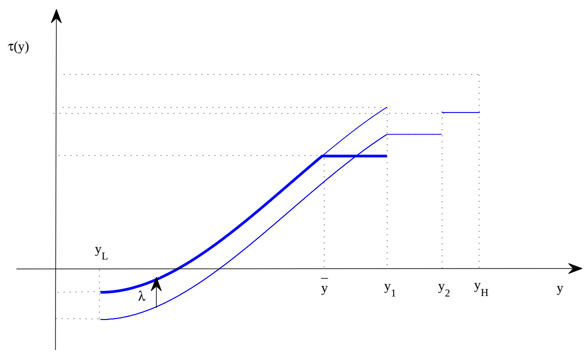

Figure 6: The Assumed Form of the Optimal Tax Function.

y_1, so initial choices in this range must remain optimal. Consequently, the amount of revenues collected from agents whose initial choices were above y_1 is not affected when $\tau_E(\,)$ is replaced with $\tau_\lambda(\,)$ defined with (123).

Agents whose initial choices were below y_1 can be split into two categories: those who do not change their behavior, for them $L_a^{**} = L_a^*$, and pay λ more in taxes, and those who change. Observe that those who change their behavior choose income levels above y_1 and as a result pay more in taxes than they initially did. Therefore, we can conclude that the government collects more in revenue if the original efficient tax function, $\tau_E(\,)$, is replaced with $\tau_\lambda(\,)$.

Now let us start a process of flattening of $\tau_\lambda(\,)$. Specifically, let us consider the following tax function

$$\tau_{\bar{y}}(y) = \begin{cases} \tau_E(y) + \lambda & \text{for } y \leq \bar{y} \\ \tau_E(\bar{y}) & \text{for } \bar{y} < y \leq y_1 \\ \tau_E(y) & \text{for } y_1 < y \end{cases}. \tag{124}$$

Observe that when $\bar{y} = y_1$ then the revenue collected with $\hat{\tau}_{\bar{y}}()$ is necessarily equal to that collected with $\tau_\lambda(\,)$ and necessarily bigger than the revenue collected with the original efficient tax function $\tau_E(\,)$. On the other hand, if $\bar{y} = y_L$ then the revenue collected must be smaller than it was initially. Invoking a form of the mean value theorem, see Appendix B for a precise statement, we can be sure that there is a value of $\bar{y}$ such that the revenue collected with $\tau_{\bar{y}}(\,)$ is the same as the revenue collected[14] with $\tau_E()$. Let us denote this value of $\bar{y}_R$. Clearly, by replacing the original efficient tax function, $\tau_E(y)$, with $\tau_{\bar{y}_R}(y)$ we preserve revenue while increasing efficiency as some agents choose to pay marginal taxes of zero while they previously did not. Therefore, we can conclude that the original tax function, $\tau_E(y)$, could not have been efficient, i.e., it could not have been a dead weight burden minimizing function.

Clearly, we have arrived at a contradiction. Our initial assumption that the efficient tax function was not of the form of a step function led us to a conclusion that the dead weight burden could be lowered, but this cannot be the case given the assumed efficiency. Consequently, we must conclude that only step functions can be viable candidates for efficient tax functions.

D Appendix

In this appendix we consider the case when the distribution of skills is unbounded, i.e., $a_H = \infty$. Note that in this case our basic procedure is not directly applicable as there is no highest level of skill, but it can be easily modified and used to argue that our results prevail even in the case when the distribution of skills is unbounded.

Let us assume that $\tau_E(\,)$ is the tax function that ensures efficiency. Again, let us consider a simple modification of $\tau_E(\,)$. In particular, assume that $\tau_E(\,)$ is shifted up by λ, i.e., we consider a function, $\tau_\lambda(\,)$, of the form $\tau_\lambda(y) = \tau_E(y) + \lambda$ for y that are actually chosen[15]. Note again that for λ sufficiently small such a shift has no bearing on individual behavior. Consequently, it is necessarily the case that the amount of revenue collected with $\tau_\lambda(\,)$ exceeds that collected with $\tau_E(\,)$. Now, let us consider simple modifications of $\tau_\lambda(\,)$. Specifically, let us consider tax function $\tau_{\bar{y}}(\,)$, which is identical to $\tau_\lambda(\,)$ for $y \leq \bar{y}$ and equal to $\tau_\lambda(\bar{y})$ everywhere else. Note that if we assume that $\bar{y} = y_L$, where y_L denotes the lowest income earned then the revenue collected with $\tau_{\bar{y}}(\,)$ must be smaller than the revenue collected with $\tau_E(\,)$ for λ small enough. As we increase $\bar{y}$ the revenue increases and ultimately it exceeds the revenue collected with $\tau_E(\,)$. Therefore, we can expect, see the discussion on Appendix B,

[14] Note that now some agents who initially chose income levels above y_1 can choose income levels below y_1 as tax liabilities in this range are smaller, which by itself causes a revenue drop in this group. This, however, does not create problems as we consider continuous variations in $\bar{y}$.

[15] We can define $\tau_\lambda(y) = y$ everywhere else.

that there exists a value of $\bar{y}$ such that it enures revenue neutrality. Let $\bar{y}_R$ be such a value. Note that once the revenue neutrality is ensured it is obvious that $\tau_{\bar{y}_R}(\,)$ becomes feasible. Moreover, by invoking similar arguments to those presented in the main body of the text we can show that the dead weight burden is never increased when $\tau_E(\,)$ is replaced with $\tau_{\bar{y}_R}(\,)$ and it is sometimes lowered. Consequently, we can state that $\tau_E()$ cannot be efficient unless it is of the form of $\tau_{\bar{y}_R}()$ to start with, which implies that our basic result that efficient tax functions must be locally flat prevails even in the case of unbounded distributions.

E Appendix

Through the paper we have kept our assumption of positive labor supply on the part of all agents. In this Appendix we show that our results hold in the case when some agents choose not to supply labor when faced with the efficient tax function $\tau_E(\,)$.

It turns out that to establish that our results prevail when some agents choose to be idle we must consider a number of options.

Case 1. Assume that $\tau_E(0) < 0$.

This case is particularly simple. If $\tau_E(0) < 0$ then of course $\tau_E(0) + \lambda < 0$ for λ small enough, i.e., $\tau_E(\,)$ shifted up by λ remains feasible. Consequently, we can can always rewrite inequalities (138) and (139) as, note that $h(\,)$ by assumption is an increasing function,

$$\forall y,\ a \in A_0|\ h(-\tau_E(0) - \lambda) \geq h(y - \tau_E(y) - \lambda - \frac{1}{2}(\frac{y}{a})^2) \tag{125}$$

and

$$\forall a \in A_+|\ h(y_a^* - \tau_E(y_a^*) - \lambda - \frac{1}{2}(\frac{y_a^*}{a})^2) \geq h(-\tau_E(0) - \lambda). \tag{126}$$

Naturally, inequalities (125) and (126) imply that agents to not have any incentives to change their behavior when $\tau_E(\,)$ is shifted up by λ. Furthermore, we can then employ our procedure of flattening the far end of $\tau_E(\,) + \lambda$ and argue that efficiency must increase if such a procedure is applied. As a result we can state that in this case, when some agents are not working, but $\tau_E(0) < 0$, our basic result stands. In other words, the efficient tax function must be locally flat.

Case 2. Assume that $\tau_E(0) = 0$ and $\tau_E^{'}(0) < 1$.

In this case we cannot directly apply our procedure as $\tau_E(0) + \lambda > 0$ for any positive λ and a simple upward shift of $\tau_E(\,)$ stops being feasible. However, given that $\tau_E^{'}(0) < 1$ we can easily[16] find a tax function of the form $\tau_L(y) = \tau y$ that lies above $\tau_E(\,)$ everywhere except for 0. Furthermore, note that agents when faced with $\tau_L(\,)$ always choose to be active. In particular, we have

$$\forall a \in [a_L, a_H]|\, y_a^L = (1-\tau)a^2 > 0. \tag{127}$$

Furthermore, preferring to be active rather than inactive requires that

$$\forall a \in [a_L, a_H]|\, h(\frac{1}{2}(1-\tau)^2 a^2) > h(0). \tag{128}$$

Furthermore, given that $\tau_L(\,)$ by construction is never below $\tau_E(\,)$ it must be the case that agents cannot be better off when they face $\tau_L(\,)$ rather than $\tau_E(\,)$, i.e., we must have[17]

$$\forall a \in [a_L, a_H]|\, h(y_a^* - \tau_E(y_a^*) - \frac{1}{2}(\frac{y_a^*}{a})^2) \geq h(\frac{1}{2}(1-\tau)^2 a^2). \tag{129}$$

Combining inequalities (128) and (129) we can establish that

$$\forall a \in [a_L, a_H]|\, h(y_a^* - \tau_E(y_a^*) - \frac{1}{2}(\frac{y_a^*}{a})^2) > h(0). \tag{130}$$

However, given that by assumption$\tau_E(0) = 0$ we can state that $h(0) = h(-\tau_E(0))$, and hence inequality (130) becomes

$$\forall a \in [a_L, a_H]|\, h(y_a^* - \tau_E(y_a^*) - \frac{1}{2}(\frac{y_a^*}{a})^2) > h(-\tau_E(0)). \tag{131}$$

However, inequality (131) implies that being active is preferred to being idle. Consequently, in this case all agents are active and our result that the efficient tax function must be locally flat, as established in the main body of the text, prevails.

Case 3. Assume that $\tau_E(0) = 0$ and $\tau_E^{'}(0) = 1$.

Let us consider here a number of sub-cases. First let us assume that there exists an agent whose productivity is equal to a and who is just indifferent between being

[16]If it ever were to happen that the tax function, $\tau_E(\,)$, actually returns and touches the 45° line again at a point different than 0 then that point and points in its neighborhood would be never chosen. See arguments applied in Case 3.

[17]Note that we are only comparing utility levels. We do not argue that $\tau_E(\,)$ and $\tau_L(\,)$ bring the same level of revenue.

active and inactive. It must, in her case, be

$$h(y_a^* - \tau_E(y_a^*) - \frac{1}{2}(\frac{y_a^*}{a})^2) = h(-\tau_E(0)) \tag{132}$$

for some y_a^*.

Naturally it is straightforward to verify that agents with lesser productivities than a choose to be inactive.

The realized utility of an agent whose productivity is equal to a is given by $h(-\tau_E(0))$. Furthermore, the agent would like to pay $T(y)$ to be able to earn y without lowering her utility. Naturally, $T(y)$ satisfies

$$T(y) = -\frac{1}{2}(\frac{y}{a})^2 + y. \tag{133}$$

Naturally, $T(y)$ is concave. Furthermore, by definition $\tau_E(\ )$ must be tangent to $T(y)$ both at 0 and y_a^* as by assumption the agent is indifferent between earning 0 and y_a^*.

Now consider the following tax function

$$\tau_M(y) = \begin{cases} min\{y,\ \tau_E'(y_a^*)(y - y_a^*) + \tau_E(y_a^*)\} & \text{for } y < y_a^* \\ \tau_E(y) & \text{for } y \geq y_a^* \end{cases}. \tag{134}$$

Now, observe that if we replace $\tau_E(\ )$ with $\tau_M(\ )$ then agents with productivities higher than a do not change their behavior as the part of the tax schedule that they chose to be on is still available to them. Furthermore, moving to the left of a is suboptimal as now it would imply higher tax liabilities than when they faced $\tau_E(\ )$. Moreover, observe that if some agents choose to be active now then they must be better off by revealed preference[18] and at the same time the revenue collected will go up, but that would put the efficiency of $\tau_E(\ )$ into question. Therefore, we must conclude that if $\tau_E()$ is replaced with $\tau_M()$ then agents remain at their original choice leaving the revenue collected and the value of the objective functional unchanged.

[18] Actually, the value of the dead weight burden declines in this case as well.

Now, let us consider the following tax function

$$\tau_{M,\tau}(y) = \begin{cases} min\{\tau y,\ \tau_E^{'}(y_a^*)(y - y_a^*) + \tau_E(y_a^*)\} & \text{for } y < y_a^* \\ \tau_E(y) & \text{for } y \geq y_a^* \end{cases} \tag{135}$$

for τ very close to 1. Again, for τ very close to 1 agents with productivities above a do not change their behavior. At the same time agents, whose productivities are below a start being active as now they effectively see τy as their tax schedule. However, if they start being active then the revenue increases and, more importantly the value of the dead weight burden declines, and, consequently, we can conclude that the original tax function $\tau_E()$ cannot be efficient. Therefore, the sub-case considered here cannot occur, i.e., our results stand.

We still have to consider one more sub-case when $\tau_E^{'}(0) = 1$, and some agents choose to be inactive, but there is no agent who is just indifferent between being active and inactive.

Let us consider the following sets

$$A_0 = \{a \in [a_L, a_H] |\ y_a^* = 0\} \tag{136}$$

and

$$A_+ = \{a \in [a_L, a_H] |\ y_a^* > 0\}. \tag{137}$$

Obviously, by revealed preference, we must always have

$$\forall y,\ a \in A_0 |\ h(-\tau_E(0)) \geq h(y - \tau_E(y) - \frac{1}{2}(\frac{y}{a})^2) \tag{138}$$

and

$$\forall a \in A_+ |\ h(y_a^* - \tau_E(y_a^*) - \frac{1}{2}(\frac{y_a^*}{a})^2) \geq h(-\tau_E(0)). \tag{139}$$

Naturally, it is straightforward to verify that if $a \in A_0$ then for any $a_1 \leq a$ it must be the case that $a_1 \in A_0$. Therefore, we can say that we must have the following

$$A_0 = [a_L, a) \text{ and } A_+ = [a, a_H] \tag{140}$$

or

$$A_0 = [a_L, a] \text{ and } A_+ = (a, a_H]. \tag{141}$$

If condition (140) holds then we can again define using a as the reference point$\tau_{M,\tau}()$ just like we did above and argue that $\tau_E(\,)$ cannot be efficient. On the other hand, if condition (141) holds then by revealed preference we must have

$$\forall y > 0|\ h(-\tau_E(0)) > h(y - \tau_E(y) - \frac{1}{2}(\frac{y}{a})^2) \tag{142}$$

and

$$\forall \varepsilon > 0|\ h(y^*_{a+\varepsilon} - \tau_E(y^*_{a+\varepsilon}) - \frac{1}{2}(\frac{y^*_{a+\varepsilon}}{a+\varepsilon})^2) > h(-\tau_E(0)). \tag{143}$$

Using simple Taylor series expansions we can reduce the above inequalities can to

$$\forall y > 0|\ 0 > \frac{1}{2}y^2(-\frac{1}{a^2} - \tau_E^{''}(\theta_y)) \tag{144}$$

and to

$$\forall \varepsilon > 0|\ \frac{1}{2}(y^*_{a+\varepsilon})^2(-\frac{1}{(a+\varepsilon)^2} - \tau_E^{''}(\theta_{y^*_{a+\varepsilon}})) > 0. \tag{145}$$

By definition y and $y^*_{a+\varepsilon}$ are strictly positive, so we can rewrite the above inequalities as

$$\forall y > 0|\ 0 > -\frac{1}{a^2} - \tau_E^{''}(\theta_y) \tag{146}$$

and

$$\forall \varepsilon > 0|\ -\frac{1}{(a+\varepsilon)^2} - \tau_E^{''}(\theta_{y^*_{a+\varepsilon}}) > 0. \tag{147}$$

Now let us denote $lim_{\varepsilon\to 0}y^*_{a+\varepsilon}$ with $y^*_{a^+}$. If $y^*_{a^+} > 0$ then inequality (146) implies

$$0 > -\frac{1}{a^2} - \tau_E^{''}(\theta_{y^*_{a^+}}). \tag{148}$$

Furthermore, letting ε approach 0 in inequality (147) allows us to establish that

$$-\frac{1}{a^2} - \tau_E^{''}(\theta_{y^*_{a^+}}) \geq 0, \tag{149}$$

which naturally contradicts inequality (148). Therefore, it must be that $y^*_{a^+} = 0$, i.e., that all value of income in any neighborhood of 0 are actually chosen by someone.

Let us consider the following tax function

$$T(y) = y - \frac{1}{2}(\frac{y}{a})^2. \tag{150}$$

Naturally, $T(y)$ denotes the amount that an agent whose productivity is equal to a would be willing to pay if she was to earn y and be as well of as she is given her true choice of $y_a^* = 0$. Note that by definition of a it must be the case that $T(y)$ is tangent to $\tau(y)$ and to the 45° line at 0. Let us consider income level $y = \varepsilon > 0$. Of course $y = \varepsilon$ is chosen by someone. Let us denote the productivity of an agent who chooses $y = \varepsilon$ with a_ε. Of course $a_\varepsilon > a$. Now, let us consider a simple modification of $\tau_E(\,)$. Specifically, let us consider the following tax function

$$\tau_M(y) = \begin{cases} min\{y,\ \tau_E'(\varepsilon)(y-\varepsilon) + \tau_E(\varepsilon) & \text{for } y \leq \varepsilon \\ \tau_E(y) & \text{for } y > \varepsilon \end{cases}. \tag{151}$$

Observe that if we replace $\tau_E(\,)$ with $\tau_M(\,)$ then agents who initially chose incomes above ε do not change their behavior. Furthermore, agents whose productivity is not higher than a and who initially were inactive remain inactive. Agents who were initially active and whose productivities were in range $(0, \varepsilon)$ remain active. This must be so given the single crossing property of the relevant indifference curves and the fact that by definition $T(y)$ and $\tau(y)$ are tangent at 0. Moreover, agents who initially chose incomes in interval $(0, \varepsilon)$ now face lower marginal tax and earn more and actually pay more in taxes. Consequently, the revenue collected increases and efficiency increases as well. This, however, implies that $\tau_E(\,)$ could not be efficient. Consequently, the case considered here cannot occur and consequently our results prevail.

Case 4. Assume that $\tau_E(0) > 0$.

This case is by assumption not possible as it violates the feasibility condition, (4).

The four cases presented above exhaust all physical possibilities. Therefore, we can conclude that our results hold even if some agents choose to be inactive.