Explaining Taxes at the Upper Tail of the Income Distribution: The Role of Utility Interdependence

Samano, Daniel

Banco de Mexico

8 December 2009

Banco de México
Documentos de Investigación

Banco de México
Working Papers

N° 2009-16

Explaining Taxes at the Upper Tail of the Income Distribution: The Role of Utility Interdependence

Daniel Sámano
Banco de México

December, 2009

Documento de Investigación
2009-16

Working Paper
2009-16

Explaining Taxes at the Upper Tail of the Income Distribution: The Role of Utility Interdependence*

Daniel Sámano†
Banco de México

Abstract: Optimal tax theory has difficulty rationalizing high marginal tax rates at the upper end of the income distribution. In this paper, I construct a model of optimal income taxation in which agents' preferences are interdependent. I derive a simple expression for optimal taxes that accommodates consumption externalities within Mirrlees (1971) framework. Using this expression, I conduct a *positive* analysis of taxation: assuming that observed taxes are optimal, I derive *analytic* expressions for i) a parameter that measures the degree of agents' utility interdependence and ii) a function that quantifies the consumption externality agents of different income impose to society. Using these expressions, I rationalize income taxes in the United States and the United Kingdom for the 1995-2004 period. I show that only a *moderate* amount of utility interdependence is sufficient for this. My estimations indicate that the progressivity of tax schedules may be driven by corrective considerations.
Keywords: optimal non-linear taxation, relative consumption, rationalization.
JEL Classification: D62, H21, H23.

Resumen: La teoría de impuestos óptimos tiene dificultad para racionalizar altas tasas marginales de impuestos al ingreso en la parte alta de la distribución. En este documento, se construye un modelo de impuestos óptimos en el cual las preferencias de los agentes son interdependientes. Se deriva una expresión simple del impuesto óptimo que acomoda externalidades de consumo dentro del marco de Mirrlees (1971). Usando esta expresión, se lleva a cabo un análisis *positivo* de impuestos: suponiendo que los impuestos observados son óptimos, se derivan expresiones *analíticas* para i) un parámetro que mide el grado de interdependencia en la utilidad de los agentes y ii) una función que cuantifica las externalidades de consumo que agentes con distintos niveles de ingreso imponen a la sociedad. Usando estas expresiones, se racionalizan los impuestos al ingreso en los Estados Unidos y en el Reino Unido en el periodo 1995-2004. Se muestra que sólo una cantidad *moderada* de interdependencia en la utilidad es suficiente para esto. Las estimaciones indican que la progresividad en los impuestos pudiera ser ocasionada por consideraciones correctivas.
Palabras Clave: impuestos óptimos no-lineales, consumo relativo, racionalización.

*I thank my advisor Narayana Kocherlakota for his guidance while at Minnesota. I am also grateful to Santiago Bazdresch, Nick Guo, Fatih Guvenen, Larry Jones, Antonio Mele, Christopher Muller, Andrew Oswald, Chris Phelan, Aldo Rustichini, Anderson Schneider, Rene Schwengber, Claudia Senik, Ina Simonovska, Adam Slawski, Matti Tuomala and participants of the Public Economics and Policy Workshop at Minnesota, the 2008 Midwest Economic Theory Meeting at Urbana-Champaign, the Relativity, Inequality and Public Policy Conference at Edinburgh, Banco de Mexico Seminar and the 2009 Royal Economic Society Conference. I am indebted to the NBER and the Economic and Social Data Service for providing the data.
† Dirección General de Investigación Económica. Email: dsamano@banxico.org.mx.

I must confess that I had expected the rigorous analysis of income-taxation in the utilitarian manner to provide an argument for high taxes. It has not done so. [Sir James A. Mirrlees, *"An Exploration in the Theory of Optimum Income Taxation"*]

1 Introduction

The trade-off between redistribution and incentives has been analyzed extensively by economists. A very important lesson extracted from these studies is that if the government redistributes excessively, highly productive individuals are left with little or no incentives to work. In order to avoid this, consumption and income inequalities arise as a consequence of incentive problems.

Mirrlees (1971) was the first study that formulated this problem in a rigorous fashion. He finds that despite inequality aversion considerations by the government, low marginal labor income taxes on affluent individuals are desirable. Other studies, such as Sadka (1976) and Seade (1977) refined one of Mirrlees (1971)'s assumptions and obtained the well known result that the tax rate at the top of the distribution must be zero. These results may be surprising, but the intuition behind them is quite clear: a high marginal income tax induces leisure of highly productive individuals. This is very costly for the economy as a whole in terms of forgone output. Thus, low marginal taxes at the high end of the income distribution are optimal despite redistributional considerations.

Having such a strong theoretical argument against, why is it that in most of the countries marginal taxes on affluent people are high?[1] One explanation is provided by Diamond (1998) and Saez (2001). They obtain positive asymptotic marginal taxes when the functional form describing the labor earnings distribution's upper tail is Pareto. However, the result no

[1] A recent article by The Economist reports that the highest marginal tax rates in Belgium, Japan and Sweden are close to 50%, in Australia, China, France, Germany and Italy are close to 45% while in Brazil, India and the United States are around 30%.

longer holds if the previous condition fails to be satisfied.[2]

In this paper, I explore an alternative, yet mutually compatible hypothesis: taxation on affluent individuals may occur in order to correct consumption externalities. To formalize this idea, I model an economy populated by a continuum of agents with heterogeneous privately-known productivities. Using a novel functional form, agents' preferences are interdependent.[3] Thus, agents' consumption generates an externality in the form of a consumption benchmark. This introduces an additional reason for government intervention, namely, taxing income for corrective purposes. This occurs in absence of a non-linear consumption tax. I derive a simple expression for optimal taxes that accommodates consumption externalities within Mirrlees (1971) framework. This expression decomposes the observed tax schedule into two components: the Mirrleesian and the Pigouvian tax. Applying this formula, I conduct a *positive* analysis of taxation: assuming that observed taxes are optimal, I derive *analytic* expressions for i) a parameter that measures the degree of agents' utility interdependence and ii) a function that quantifies the consumption externality agents of different income impose to society. Using these expressions, I calculate the magnitude of consumption externalities that rationalize labor income taxes in the United States and the United Kingdom from 1995 to 2004. I show that only a *moderate* amount of what is known in the literature as "jealousy" (see Dupor and Liu (2003)) toward affluent individuals is sufficient to rationalize the observed labor income taxes in the United States and the United Kingdom for the aforementioned period. This result suggest that the progressivity of actual tax schedules may be driven by corrective considerations, particularly at the top of the earnings distribution. This is the main contribution of this paper.

If actual fiscal policy is supposed to be influenced by positional concerns, we need to gather empirical evidence in this regard. In this respect, there are

[2] I am explicit about this condition later in the text.

[3] A non-exhaustive list of economists who supported the idea that relative rather than absolute consumption matters are Veblen (1899) and Duesenberry (1949) .

a few studies that have attempted to measure the degree to which relative consumption matters for people's satisfaction. Surveying individuals about their choice among hypothetical worlds they could live in is one approach.[4] In world A, the assets of the subject are higher than in world B. However, agents are worse off in world A than in world B with respect to the population average. Thus, individuals' choices end up revealing their concern for relative positions. For the sake of concreteness, most surveys focus on particular assets such as cars, houses and leisure. In some cases, they also include income. Using a survey applied to a representative sample of the Swedish population, Carlsson, Johansson-Stenman, and Martinsson (2007) find evidence that supports the relative consumption hypothesis for income and cars but not for leisure.[5] Alpizar, Carlsson, and Johansson-Stenman (2005) survey students from Costa Rica and obtain similar results. J. Solnick and Hemenway (1998) survey a sample of American students and find that about 50% of them would prefer a world in which they had half their absolute income as long as their relative standing was high.

An alternative and more common approach for testing utility interdependence is applying a regression analysis. Using data on British workers job satisfaction, Clark and Oswald (1996) construct reference groups that comprise individuals with the same labor market characteristics such as age, education, sex, monthly earnings and hour per week worked. They find that people are less satisfied with their jobs the higher the income of their reference group is. Luttmer (2005) merges a database on individuals' self reported happiness to information about local (geographically speaking) average earnings and finds that self reported happiness is negatively affected by the earnings of others in their area. A non-exhaustive list of other pa-

[4] To be more precise, most of those surveys ask individuals where they would like an imagined future relative of them to live in. According to Alpizar, Carlsson, and Johansson-Stenman (2005), this in order to liberate them from current circumstances.

[5] Some of the results of this paper are striking. For instance, they find that about 50% of the utility obtained from cars and income comes from relative concerns.

pers that have tested the significance of others' consumption or income on individuals' well-being are McBride (2001), Ferrer-i Carbonell (2005), Dynan and Ravina (2007), Senik (2008) and Clark and Senik (2008). On social comparisons among family members, Neumark and Postlewaite (1998) find that a woman outside the formal labor force is 16-25% more likely to work outside the home if her sister's husband earns more than her own husband. At an experimental level, Rustichini and Vostroknutov (2007) conduct a "burning money" game and find that individuals are willing to incur a cost in order to reduce the winnings of others. This occurs mostly when the winnings result from skill rather than luck. Bault, Coricelli, and Rustichini (2008) conduct an experiment in which participants choose among lotteries with different levels of risk, and can observe the choice that others have made. Based on subjective emotional evaluations and physiological responses, they find that individuals experience jealousy and gloating upon comparing their outcomes.

Regarding studies of optimal taxation under consumption externalities in dynamic settings, the work of Ljungqvist and Uhlig (2000) presents a complete markets dynamic economy driven by productivity shocks. The negative externality they model is known as external habit formation.[6] That is, when agents increase their consumption, they do not take into account their effect over the aggregate desire of other agents to catch up. Nevertheless, since the consumption distribution is degenerate in their model, the policy implications that can be extracted are to prevent consumption addiction, not jealousy itself. Abel (2005) models an overlapping generations economy in which agents display jealousy toward the consumption of all living generations at a given period. When the social planner is more patient than individuals, he finds that it is optimal to tax capital in order to transfer consumption from old agents to young ones. This is true since the planner wants to reallocate consumption toward later generations of consumers.

[6]External habit formation was first introduce in the finance literature in order to explain the equity premium puzzle. See Abel (1990), Constantinides (1990), Gali (1994), Heaton (1995) and Campbell and Cochrane (1999).

To the best of my knowledge, the only studies prior to mine regarding optimal income taxation à la Mirrlees under utility interdependence are Oswald (1983) and Tuomala (1990).[7] Both articles conduct a *normative* analysis of taxation and characterize neatly optimal tax rules under the assumption that agents value their consumption relative to the average consumption. Oswald (1983) highlights that the zero marginal tax rate at the extremes result no longer holds when agents' preferences are interdependent. This article also points out that higher taxes are optimal in a predominantly jealous world while the opposite is true in an economy populated mostly by altruistic agents. The numerical calculations of Tuomala (1990) also show that optimal income taxes are progressive and that higher overall taxes correspond to a higher degree of jealousy in agents' preferences. In contrast to my work, neither article attempts to rationalize observed tax schedules. Another related paper is Ireland (2001) which introduces wasteful consumption through which individuals signal their skills. The main result is that the optimal nonlinear tax schedule in this environment is higher but not more progressive than without status seeking. Moreover, when the distribution of skills is bounded, the zero marginal tax rates at the extremes result is unaffected by the signaling mechanism.

The rest of this paper proceeds as follows: section 2 presents the model, section 3 describes the data and estimation procedure, section 4 shows and discusses the results and section 5 concludes.

[7] See also Boskin and Sheshinski (1978) for a classical study of *affine* income taxation when individuals value their consumption with respect to the average consumption.

2 The Model

Consider a static economy populated by a continuum of agents with heterogeneous productivity or skill. Let $\theta \in \Theta$, where $\Theta \equiv [\underline{\theta}, \bar{\theta}]$ and $0 < \underline{\theta} < \bar{\theta} < \infty$, be individual's productivity distributed according to the density $f : \Theta \to \mathbb{R}_{++}$. Productivity is *privately* known to each agent. An agent with productivity θ has a utility function separable in "consumption" and leisure of the form

$$U(c, y, C; \theta) = u(c, C) - v\left(\frac{y}{\theta}\right)$$

where c, y are individual's consumption and effective labor, respectively.[8] Moreover,

$$C \equiv \int_{\Theta} c(\theta)\psi(\theta)d\theta \tag{1}$$

is the society's consumption benchmark specified as a *weighted* average of consumption. As usual, preferences satisfy $u_c > 0$, $u_{cc} \leq 0$, $v' > 0$ and $v'' > 0$, and $u(\cdot)$ is jointly concave.[9] I also impose the condition that $u_C < 0$.

According to the previous specification, individuals value their own consumption *relative* to what the rest of all individuals consume. Hence, the so called "reference group" in this economy is the whole society itself. Since the utility of an agent decreases as the weighted average of consumption increases, we say that preferences exhibit what is known in the literature as "jealousy". This is in line with the terminology proposed in Dupor and Liu (2003). It is important to remark that according to (1), the weighting function $\psi(\theta) : \Theta \to \mathbb{R}$ does not need to be equal to $f(\theta)$.[10] In other words,

[8] As standard in this literature, I define effective labor as $y = \theta l$ where l is the amount of time worked.

[9] Utility specification such as $u(c, C) = \tilde{u}(c - \alpha C)$, $\alpha \in [0, 1]$ with $\tilde{u}' < 0$ and $\tilde{u}'' \leq 0$ satisfy those assumptions.

[10] Most of the literature on relative consumption valuation assumes that individuals value their own consumption relative to the *average* consumption. In this case $\psi(\theta) = f(\theta)$. Examples of this are Oswald (1983) and Tuomala (1990). Abel (2005) is an exception since this paper models an overlapping generations economy whose consumption benchmark is

agents may contribute to the consumption externality that society faces in a magnitude different from their population size. Hereafter, I will refer to $\psi(\theta)$ as the externality weighting function. For further reference, notice that equation (1) can be reexpressed as

$$C \equiv \int_{\Theta} c(\theta) f(\theta) \frac{\psi(\theta)}{f(\theta)} d\theta \tag{2}$$

hence the ratio $\frac{\psi(\theta)}{f(\theta)}$ acts as a weighting variable of the consumption externality that the society faces.

An *allocation* in this economy is $\{c(\theta), y(\theta)\}_{\theta \in \Theta}$, where $c : \Theta \to \mathbb{R}_+$ and $y : \Theta \to \mathbb{R}_+$. Abstracting from government expenditure, I define an allocation $\{c(\theta), y(\theta)\}_{\theta \in \Theta}$ to be *feasible* if

$$\int_{\Theta} c(\theta) f(\theta) d\theta = \int_{\Theta} y(\theta) f(\theta) d\theta \tag{3}$$

Making use of the Revelation Principle, an allocation is *incentive compatible* if

$$u(c(\theta), C) - v\left(\frac{y(\theta)}{\theta}\right) \geq u(c(\theta'), C) - v\left(\frac{y(\theta')}{\theta}\right) \quad \forall \theta, \theta' \in \Theta \tag{4}$$

Observe that since C cannot be affected unilaterally by a single agent, it is not a function of θ. An allocation that is incentive compatible and feasible is said to be *incentive-feasible*. Finally, let $g : \Theta \to \mathbb{R}_+$ be the density according to which individuals are weighted by the benevolent planner.

a weighted geometric average of the consumption of individuals belonging to two different generations.

Definition 1. *An optimal allocation is an allocation* $\{c^*(\theta), y^*(\theta)\}_{\theta\in\Theta}$ *that maximizes the following planner problem*

$$\int_\Theta \left[u(c(\theta), C) - v\left(\frac{y(\theta)}{\theta}\right)\right] g(\theta)d\theta \tag{5}$$

subject to $\{c(\theta), y(\theta)\}_{\theta\in\Theta}$ *being incentive-feasible and* C *as defined in (1).*

2.1 Characterization of Optimal Allocations

The following proposition states the necessary conditions that any interior optimal allocation must satisfy. Let $\epsilon^*(\theta) \equiv \frac{v'(\frac{y^*(\theta)}{\theta})}{v''(\frac{y^*(\theta)}{\theta})\frac{y^*(\theta)}{\theta}}$.

Proposition 1. *Any interior optimal allocation* $\{c^*(\theta), y^*(\theta)\}_{\theta\in\Theta}$ *must be incentive-feasible and satisfy*

$$\frac{u_c(c^*(\theta), C^*)}{v'(\frac{y^*(\theta)}{\theta})\frac{1}{\theta}} - 1 = \frac{\gamma\psi(\theta)}{\lambda f(\theta)} + \frac{u_c(c^*(\theta), C^*)}{\theta f(\theta)}\left[1 + \frac{1}{\epsilon^*(\theta)}\right] \times$$

$$\int_{\underline{\theta}}^{\theta} \left[\frac{g(t)}{\lambda} - \frac{f(t)}{u_c(c^*(t), C^*)} - \frac{\gamma}{\lambda}\frac{\psi(t)}{u_c(c^*(t), C^*)}\right] dt \tag{6}$$

$$\frac{\gamma}{\lambda} = \frac{-\int_\Theta \frac{u_C(c^*(\theta),C^*)}{u_c(c^*(\theta),C^*)} f(\theta)d\theta}{1 + \int_\Theta \frac{u_C(c^*(\theta),C^*)}{u_c(c^*(\theta),C^*)}\psi(\theta)d\theta}. \tag{7}$$

where

$$C^* \equiv \int_\Theta c^*(\theta)\psi(\theta)d\theta \tag{8}$$

Proof. See Appendix A. □

Notice that the previous solution collapses into the solution of a Mirrlessian economy with no consumption externalities when the Lagrange multiplier associated to equation (1) in the maximization problem, γ, equals zero. This is true since in that case $u_C(c(\theta), C) = 0$. In section 2.4, I make assumptions that facilitate the interpretation of the previous expression.

2.2 Implementation

Agents in this economy trade effective labor for consumption. There is a single firm that employs all agents. It produces one unit of output for every unit of effective labor, y. Every unit of effective labor receives a payment of w. Agents are also subject to an income tax schedule $T(y(\theta))$, assumed to be twice differentiable and to induce no bunching. Without loss of generality, there are not taxes on consumption c. An agent with effective labor y pays taxes $T(y(\theta))$. Thus, taking as given $T(y(\theta))$, C and the wage w, the problem solved by the agent with productivity $\theta, \forall\theta \in \Theta$ is

$$\max_{c(\theta),y(\theta)} u(c(\theta), C) - v\left(\frac{y(\theta)}{\theta}\right) \tag{9}$$

s.t.

$$c(\theta) \leq wy(\theta) - T(y(\theta))$$

$$c(\theta), y(\theta) \geq 0$$

Definition 2. *Given a labor tax $T(y(\theta))$ and C, an equilibrium in this economy is an allocation $\{c^{eq}(\theta), y^{eq}(\theta)\}_{\theta\in\Theta}$ and wage w^{eq} such that*
i. $(c^{eq}(\theta), y^{eq}(\theta))$ *solve (9)* $\forall\theta \in \Theta$
ii. $C = \int_\Theta c^{eq}(\theta)\psi(\theta)d\theta$
iii. $w^{eq} = 1$
iv. $\int_\Theta T(y^{eq}(\theta))f(\theta)d\theta = 0$
v. $\int_\Theta c^{eq}(\theta)f(\theta)d\theta = \int_\Theta y^{eq}(\theta)f(\theta)d\theta$

An allocation $\{c(\theta), y(\theta)\}_{\theta\in\Theta}$ is *implementable* by the income tax $T(y(\theta))$ if $\{c(\theta), y(\theta)\}_{\theta\in\Theta}$ and w are an equilibrium.

2.3 Characterization of Optimal Income Tax

Define the following tax mechanism $T : y \to \mathbb{R}$,

$$T(y(\theta)) = \begin{cases} y^*(\theta) - c^*(\theta) & \text{if } y(\theta) = y^*(\theta) \\ y(\theta) & \text{otherwise.} \end{cases} \tag{10}$$

together with

$$\frac{T'(y^*(\theta))}{1 - T'(y^*(\theta))} = \frac{\gamma\psi(\theta)}{\lambda f(\theta)} + \frac{u_c(c^*(\theta), C^*)}{\theta f(\theta)}\left[1 + \frac{1}{\epsilon^*(\theta)}\right] \times$$

$$\int_{\underline{\theta}}^{\theta}\left[\frac{g(t)}{\lambda} - \frac{f(t)}{u_c(c^*(t), C^*)} - \frac{\gamma}{\lambda}\frac{\psi(t)}{u_c(c^*(t), C^*)}\right] dt \tag{11}$$

if $y(\theta) = y^*(\theta)$ and where $\frac{\gamma}{\lambda}$ follows expression (7).

Proposition 2. *Any optimal allocation $\{c^*(\theta), y^*(\theta)\}$ can be implemented by a tax schedule $T(y(\theta))$ defined by (10) and (11).*

Proof. See Appendix A. □

2.4 The Quasi-linear Environment

In order to keep the model tractable and facilitate the interpretation of the optimal income tax, I consider the case of

$$u(c, C) = (1-\alpha)c + \alpha(c - C), \alpha \in [0, 1), \tag{12}$$

where C is defined as in (1).[11] There are two important features to highlight about this preference specification.[12] First, the parameter α measures positional concerns agents may have. As α approaches one from below, an individual is almost as well off consuming an extra unit or observing the consumption of all other agents being reduced by the same amount. Conversely, when $\alpha = 0$, the consumption of others is completely irrelevant for an individual's satisfaction. Second, under the specification in (12), the shadow price of aggregate consumption is $1 - \alpha$. That is, if all agents in an economy were to consume one unit of the good, a share α of aggregate utility would vanish due to the jealousy effect. To see why, suppose all agents were to be given one unit of the consumption good. Such consumption would provide only $1-\alpha$ utils to an agent after she realizes that not just her, but *all* agents are consuming an extra unit. Hereafter, I will refer to the term α as the jealousy parameter.

To state the next proposition, let $G(\theta) \equiv \int_{\underline{\theta}}^{\theta} g(t)dt$, $F(\theta) \equiv \int_{\underline{\theta}}^{\theta} f(t)dt$ and $\Psi(\theta) \equiv \int_{\underline{\theta}}^{\theta} \psi(t)dt$.

[11] Notice that this expression is equivalent to $u(c, C) = c - \alpha C$. Carlsson, Johansson-Stenman, and Martinsson (2007) use a very similar functional form. The difference is that for them, C is the average consumption. They use this functional form to measure what they call *marginal degree of positionality*. It is the fraction of the marginal utility in income that is due to the increase in relative income, the term, $c - C$. For this specification, the marginal degree of positionality is α.

[12] See Hopkins (2008) for an excellent survey on theoretical models of relative concerns and their relation to inequality.

Proposition 3. *Suppose $u(c,C) = c - \alpha C, \alpha \in [0,1)$ and C defined as in (1), then any optimal marginal income tax satisfies*

$$\frac{T'(y^*(\theta))}{1-T'(y^*(\theta))} = \underbrace{\frac{[1+\epsilon^*(\theta)^{-1}]}{\theta f(\theta)}[G(\theta)-F(\theta)]}_{\textit{Mirrleesian tax}} +$$

$$\underbrace{\frac{\alpha}{1-\alpha}\left[\frac{\psi(\theta)}{f(\theta)} + \frac{[1+\epsilon^*(\theta)^{-1}]}{\theta f(\theta)}[G(\theta)-\Psi(\theta)]\right]}_{\textit{Pigouvian tax}} \tag{13}$$

Proof. See Appendix A □

From (13), it is easy to see that without utility interdependence ($\alpha = 0$), the optimal income tax is simply the Mirrleesian tax.[13] Moreover, observe that if $g(\theta) = f(\theta) \forall \theta \in \Theta$, the Mirrleesian tax is equal to zero. Atkinson (1990) refers to this case as complete distributional indifference and it arises as the marginal utility is constant across agents regardless their consumption level.

Regarding the Pigouvian tax, notice that it is increasing in the parameter α. The term $\frac{\psi(\theta)}{f(\theta)}$ affects this tax component since this is precisely the magnitude that the Pigouvian tax corrects directly. But, what is the role of the term $G(\theta) - \Psi(\theta)$? If the planner redistributive taste is such that $G(\theta) \leq \Psi(\theta) \forall \theta \in \Theta$, then the Pigouvian tax will be reduced over its *direct* component, the term $\frac{\alpha}{1-\alpha}\frac{\psi(\theta)}{f(\theta)}$. Conversely, if $G(\theta) \geq \Psi(\theta) \forall \theta \in \Theta$ then the Pigouvian tax is higher.

Corollary 1 (Proposition 3)**.** *Marginal taxes at the top and the bottom of the earnings distribution satisfy* $\frac{T'(y^*(\bar{\theta}))}{1-T'(y^*(\bar{\theta}))} = \frac{\alpha}{(1-\alpha)}\frac{\psi(\bar{\theta})}{f(\bar{\theta})}$ *and* $\frac{T'(y^*(\underline{\theta}))}{1-T'(y^*(\underline{\theta}))} = \frac{\alpha}{(1-\alpha)}\frac{\psi(\underline{\theta})}{f(\underline{\theta})}$.

This corollary makes clear that non-zero taxation at the top and the bottom of the earnings distribution is optimal whenever $\alpha > 0$. This is due

[13] See Salanié (1997) for a direct derivation of the Mirrleesian tax for a quasi-linear model without consumption externalities. The quasi-linear case has also been analyzed by Atkinson (1990) and Diamond (1998).

to corrective considerations. This feature of the optimal income tax when preferences exhibit utility interdependence has been emphasize before by Oswald (1983) and Tuomala (1990). Moreover, observe that unless $\psi(\theta) = f(\theta) \forall \theta \in \Theta$, taxes at the top and the bottom do not need to be equal. These features of the optimal income tax occur despite the Mirrleesian component at the extremes being zero.

In Appendix B, I derive an expression for the asymptotic optimal income tax when $f(\theta)$ is distributed according to a Pareto distribution. A similar case (without consumption externalities) is analyzed by Diamond (1998) and Saez (2001) based on the premise that the upper tail of the earnings distribution can be well approximated by the aforementioned distribution. Intuitively, the upper tail of the earnings distribution is thick. Such insight, however, may need to be taken with caution after analyzing the very top of the income distribution in the U.S. for several years using non-parametric smoothing techniques. I show that the ratio $\frac{1-F_Y(y)}{yf_Y(y)}$, where y is labor income, $f_Y(y)$ is the density and $F_Y(y)$ is the earnings c.d.f., has a decreasing shape at very high labor income leves. This analysis is also presented in Appendix B. This implies that the thickness of the empirical earnings distribution may not be enough to explain positive taxes at the high end of the labor income distribution as in Mirrlees (1971). Nonetheless, as seen in Corollary 1, positional concerns and corrective considerations can account for that.

2.5 Recovering the Externality Weighting Function and the Jealousy Parameter

In this section I state and prove my main theoretical result. Put simply, this result states that conditional on a social planner's weighting function $g(\theta)$ and given a *marginal* income tax schedule $T'(y)$, gross income density $f_Y(y)$ and labor supply elasticity, it is always possible to find the externality weighting function $\psi(\theta)$ and the jealousy parameter α that rationalize the observed marginal income tax schedule. Given the imposed functional forms, the strength of this result is the possibility it creates to apply the model to the data.

Definition 3. *The parameter $\alpha \in \mathbb{R}$ and the externality weighting function $\psi : \Theta \to \mathbb{R}$ rationalize a marginal tax schedule $T'(y)$ if the resulting equilibrium allocation $\{c^{eq}(\theta), y^{eq}(\theta)\}_{\theta\in\Theta} = \{c^*(\theta), y^*(\theta)\}_{\theta\in\Theta}$.*

Theorem 1. *Suppose $T'(y) \in \mathcal{C}$ $\forall y \in [\underline{y}, \bar{y}]$ with $\underline{y} > \max\{0, \tilde{y}\}$, $\tilde{y} = \inf\{y \mid (1 - T'(y))\phi y^{\phi-1} + T''^{\phi} > 0\}$. If $u(c, C) = c - \alpha C$, $C = \int_\Theta c(\theta)\psi(\theta)d\theta$, $v\left(\frac{y}{\theta}\right) = \frac{1}{1+\phi}\left(\frac{y}{\theta}\right)^{1+\phi}, \phi > 0$, then there exists a unique mapping*

$$\mathcal{M} : (g(\cdot); \phi, f_Y(y), T'(y)) \to (\psi(\cdot), \alpha)$$

that rationalizes $T'(y)$. The externality weighting function $\psi : \Theta \to \mathbb{R}$ and the jealousy parameter $\alpha \in \mathbb{R}$ that rationalize $T'(y)$ satisfy

$$\psi(\theta) = b(\theta) + (1 + \phi)\theta^{\phi} \left[\int_{\underline{\theta}}^{\theta} \frac{b(t)}{t^{1+\phi}} dt\right] \tag{14}$$

where

$$b(\theta) \equiv \frac{(1 - \alpha)f(\theta)T'(y(\theta))}{\alpha(1 - T'(y(\theta)))} - \frac{(1 + \phi)G(\theta)}{\alpha\theta} + \frac{(1 - \alpha)(1 + \phi)F(\theta)}{\alpha\theta} \tag{15}$$

$$\alpha = \frac{\int_\Theta \frac{f(\theta)}{\theta^{\phi+1}(1-T'(y(\theta)))} d\theta - \int_\Theta \frac{g(\theta)}{\theta^{1+\phi}} d\theta}{\int_\Theta \frac{f(\theta)}{\theta^{\phi+1}(1-T'(y(\theta)))} d\theta} \tag{16}$$

and

$$f(\theta) = f_Y(\Phi^{-1}(\theta)) \frac{\partial \Phi^{-1}(\theta)}{\partial\theta} \tag{17}$$

where $\Phi(y) = \left[\frac{y^{\phi}}{1-T'(y)}\right]^{\frac{1}{1+\phi}}$. If in addition, $\int_\Theta \frac{f(\theta)}{\theta^{\phi+1}(1-T'(y(\theta)))} d\theta - \int_\Theta \frac{g(\theta)}{\theta^{1+\phi}} d\theta \geq 0$, then $\alpha \in [0, 1)$.

Proof. If $u(c, C) = c - \alpha C$ and $v(\frac{y}{\theta}) = \frac{1}{1+\phi}(\frac{y}{\theta})^{1+\phi}$, it follows from Proposition

3 that

$$\frac{T'(y(\theta))}{1-T'(y(\theta))} = \frac{\alpha}{(1-\alpha)}\frac{\psi(\theta)}{f(\theta)} + \frac{(1+\phi)}{(1-\alpha)\theta f(\theta)}\left[G(\theta)-(1-\alpha)F(\theta)-\alpha\Psi(\theta)\right] \tag{18}$$

where $\psi(\theta) = \Psi'(\theta)$. Expression (18) is a first order ordinary differential equation of the form

$$\psi(\theta) + a(\theta)\Psi(\theta) = b(\theta)$$

where $a(\theta) \equiv -\frac{1}{\theta}[1+\phi]$ and $b(\theta) \equiv \frac{(1-\alpha)f(\theta)T'(y(\theta))}{\alpha(1-T'(y(\theta)))} - \frac{(1+\phi)G(\theta)}{\alpha\theta} + \frac{(1-\alpha)(1+\phi)F(\theta)}{\alpha\theta}$. If $a(\theta)$ and $b(\theta)$ are continuous on Θ, according to Coddington (1989), Theorem 3, Chapter 1, $\Psi(\theta)$ satisfies

$$\Psi(\theta) = e^{-\int_{\underline{\theta}}^{\theta} a(t)dt}\left[\int_{\underline{\theta}}^{\theta} e^{\int_{\underline{\theta}}^{t} a(x)dx} b(t)dt + \kappa\right] \tag{19}$$

The fact that $T'(y)$ is continuous $\forall y \in [\underline{y}, \bar{y}]$ with $\underline{y} > 0$ guarantee the continuity of $a(\theta)$ and $b(\theta)$. By setting $\kappa = 0$ in (19) we have $\Psi(\underline{\theta}) = 0$ since $\Psi(\theta)$ is a cumulative weighting function. Moreover, since α is an argument of $\Psi(\theta)$ in (19) through $b(t)$, setting

$$\alpha = \frac{\int_{\Theta} \frac{f(\theta)}{\theta^{\phi+1}(1-T'(y(\theta)))}d\theta - \int_{\Theta} \frac{g(\theta)}{\theta^{1+\phi}}d\theta}{\int_{\Theta} \frac{f(\theta)}{\theta^{\phi+1}(1-T'(y(\theta)))}d\theta}$$

we obtain $\Psi(\bar{\theta}) = 1$. Using $\int_{\Theta} \frac{f(\theta)}{\theta^{\phi+1}(1-T'(y(\theta)))}d\theta - \int_{\Theta} \frac{g(\theta)}{\theta^{1+\phi}}d\theta \geq 0$, we get $\alpha \in [0, 1)$.

The last step involves the identification of the skills distribution $f(\theta)$ from income distribution $f_Y(y)$. For that, notice that from the first order condition of the consumer's problem we have $(1 - T'(y)) = \left(\frac{y}{\theta}\right)^{\phi}\left(\frac{1}{\theta}\right)$ and hence, $\theta = \Phi(y) = \left[\frac{y^{\phi}}{1-T'(y)}\right]^{\frac{1}{1+\phi}}$. Therefore, $f(\theta) = f_Y(\Phi^{-1}(\theta))\frac{\partial\Phi^{-1}(\theta)}{\partial\theta}$. The fact that $y > \tilde{y}$ guarantees the invertibility of $\Phi(\cdot)$. □

In order to gain insight into how expression (16) makes possible calculating the jealousy parameter, α, let us consider a very simple example. Suppose that $g(\theta) = f(\theta) \forall \theta \in \Theta$. Thus, we are in a case of distributional indifference and as established before, the observed marginal tax $T'(y(\theta))$ must be purely Pigouvian (see equation (13)). Further, let us assume that $T'(y(\theta)) = T' \forall y \in [\underline{y}, \bar{y}]$, i.e., society faces a flat tax. Evaluating equation (16), we obtain $\alpha = T'$. Thus, we recover the jealousy parameter from taxes. We can then proceed to calculate $\psi(\theta)$, the externality weighting function, using equation (14).

2.6 An Upper Bound for the Jealousy Parameter and a Lower Bound for the Externality at the Top

As stated in Theorem 1, it is possible to calculate $\psi(\theta)$ and α conditional on the redistributive taste of the planner represented by the density $g(\theta)$. This element, however, is non observable and thus the model I present has an identification problem. Nevertheless, it is possible to calculate an upper bound of α under the loose assumption that society favors some redistribution. This statement is formally expressed in Proposition 4.

Proposition 4. *Let* $\Gamma_F \equiv \{g : \Theta \to \mathbb{R}_+ \mid G(\theta) \geq F(\theta) \quad \forall \theta \in \Theta\}$. *Then* $\alpha \leq \bar{\alpha} \quad \forall g \in \Gamma_F$, *where* α *satisfies (16) and*

$$\bar{\alpha} = \frac{\int_\Theta \frac{f(\theta)T'(y(\theta))}{\theta^{1+\phi}(1-T'(y(\theta)))} d\theta}{\int_\Theta \frac{f(\theta)}{\theta^{1+\phi}(1-T'(y(\theta)))} d\theta} = \frac{E_Y[y^{-\phi}T'(y)]}{E_Y[y^{-\phi}]}$$

Moreover,

$$f(\theta) = \arg\max_{g(\cdot)} \{\alpha \mid g \in \Gamma_F\}$$

Proof. See Appendix C. □

Proposition 5. *Suppose* $T'(y(\bar{\theta})) \geq T'(y(\theta)) \forall \theta \in \Theta$ *with strict inequality for some interval* $I \equiv [a, b] \subset \Theta$, *with* $a \geq \underline{\theta}$, $b < \bar{\theta}$ *then* $\frac{\psi(\bar{\theta})}{f(\bar{\theta})} > 1 \; \forall g \in \Gamma_F$.

Proof. See Appendix C. □

Two things need to be highlighted about Proposition 4. First, the calculation of the upper bound of the jealousy parameter α requires only observed variables: the labor earnings distribution, the marginal tax schedule on labor income and the elasticity of labor supply. Second, the upper bound of the jealousy parameter is attained when the planner is utilitarian, i.e. $g(\theta) = f(\theta) \forall \theta \in \Theta$. The intuition for this result is straightforward: when the planner is utilitarian, there is complete distributional indifference (given the quasi-linearity in preferences) and the Mirrleesian tax is zero at any income level. Thus, all taxation must be Pigouvian. Moreover, under the assumption that societies redistribute, Proposition 5 finds a *lower* bound for the contribution to the consumption externality at the top of the income distribution, the ratio $\frac{\psi(\bar{\theta})}{f(\bar{\theta})}$. In other words, the weight by which the agent at the top of the earnings distribution contributes to the consumption externality must be strictly greater than one if observed income taxes are thought to be optimal, partly due to corrective considerations. This implies that the average consumption may not be an accurate approximation of the consumption reference point a society faces but rather a *weighted* average where the top is weighted more. In the next section, I estimate the upper bound of the jealousy parameter, $\overline{\alpha}$, and the corresponding externality weighting function $\psi(\cdot)$ for the U.S. and the U.K.

3 The Data

My numerical analysis was done for the U.S. and the U.K. This choice was based on the public availability of micro-file tax data that I describe in this section. For the U.S, I use the Statistics of Income (SOI) Public Use Tax files elaborated by the Internal Revenue Service (IRS) and distributed by the NBER.[14] The data consists of the information that U.S. citizens and

[14] See Internal-Revenue-Service (1995-2004).

residents submit to the IRS through the 1040, 1040A and 1040EZ tax forms. Cross-section samples of approximately 100,000 to 150,000 observations are available for each year from 1960 to 2004. Such samples were designed to make national level estimates by including a weighting variable to make up for the stratified nature of the sample.[15] In order to abstract from capital holdings, the definition of *gross* income that I use is salaries and wages. This is the entry of the 1040 tax forms specified as "Wages, salaries, tips, etc". In addition, I use the the marginal income tax corresponding to different income brackets as a proxy of the labor tax schedule. The marginal income tax was collected from the "Tax Rate Schedule" from 1995 to 2004 published by the IRS.

For the U.K, I use the Survey of Personal Incomes (SPI) Public Use File from the Economic and Social Data Service provided by the University of Essex.[16] These files are compiled by Her Majesty's Revenue and Customs: Knowledge, Analysis & Intelligence, and are based on information held by HM Revenue and Customs (HMRC) tax offices on individuals who could be liable for U.K. taxes. Cross-section samples of approximately 450,000 observations are available for each year from 1995 to 2004.[17] The data set also contains a variable that allows to obtain figures for the whole U.K. population. I use the variable defined as *pay* that stands for before taxes pay from employment net of benefits and foreign earnings. Marginal income taxes are collected from the "Survey of Personal Incomes Public Use Tape Documentation: Annex D: Rates of Income Tax: 1990-91 to 2004-05" located in HM-Revenue-&-Customs (1998-2007). This is the variable that I use as a

[15] The General Description Booklet for the Public Use Tax Files (several years) indicates that the sample design is a stratified probability sample and the population of tax return is classified into sub-populations (strata). According to the same source, independent samples are selected independently from each stratum. A weighting variable is obtained by dividing the population count of returns in a stratum by the number of sample returns for that stratum.

[16] See HM-Revenue-&-Customs (1998-2007).

[17] There is an extra file for the fiscal year 1985-1986. Moreover, the HM Revenue and Customs: Knowledge, Analysis & Intelligence recently released the file for 2005.

proxy of the labor tax schedule for the U.K.

3.1 Estimation of Gross Labor Income Densities

The period of analysis is 1995-2004 since I have comparable data for the two countries during these years. Even though the model is static, taking into account the gross income distributions over several years allows me to obtain a more "robust" distribution for each country calculated as

$$f_Y(y) = \frac{1}{10} \sum_{t=1995}^{2004} f_{Y,t}(y) \tag{20}$$

To calculate $f_{Y,t}(y)$ for $t = 1995, ..., 2004$, I selected the domain $y > \$100$ measured in 2004 dollars. I calculated $f_{Y,t}(y)$ using a gaussian kernel over many points unequally separated, locating more points at the bottom of the domain than at the top. This was done in order to obtain more accurate estimates from numerical integration while at the same time keeping moderate the number of grid points.[18] Income observations were weighted to obtain population estimates.

Since income observations are more sparse as income is higher, I smoothed the data after transforming it into a logarithmic scale. According to Wand, Marron, and Ruppert (1991), this transformation is appropriate under the presence of a global width parameter h and data being more sparse at the top than at the bottom of its domain. Without this transformation, either the data at the bottom of the distribution would be over-smoothed or the tail would exhibit spurious bumps. The smoothing window or width for the U.S. was set at $h = 0.9 \times s.d. \times n^{-1/5}$, where n is the number of observations and $s.d.$ is the standard deviation of $\log(y_i)$, for $i = 1, ..., n$.[19] Silverman

[18] For both countries, the number of grid points was 5,000. Moreover, I also performed my calculations using alternative kernel functions such as the Epanechnikov and Triangular and results were almost identical. The latter was not surprising given the sample size. See Silverman (1986). I use the trapezoid method for the numerical integration.

[19] Indeed, the formula suggested by Silverman (1986) is $width = 0.9 \times A \times n^{-1/5}$, where

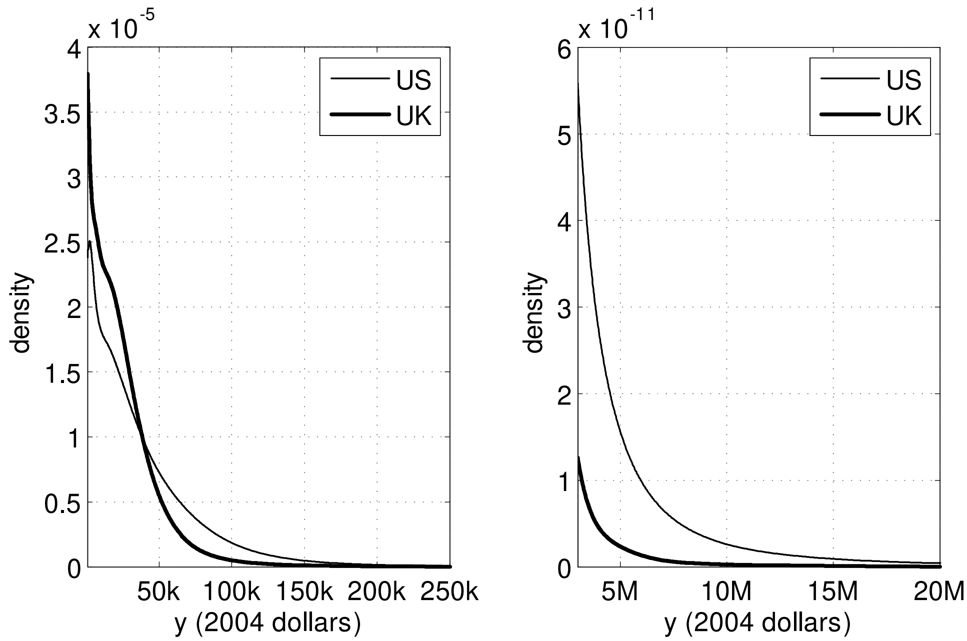

FIGURE 1: Gross Labor Income Distribution in the U.S. and the U.K. 1995-2004

(1986) indicates that such a choice of width performs very well in terms of mean integrated squared error for a wide range of densities. For the U.K., I set $h = 2 \times s.d. \times n^{-1/5}$ since lower widths produced a bump at the top of the distribution.[20] Figure 1 shows the estimated densities. I follow the convention to abbreviate 1000 dollars as k and a million dollars as M.

$A = \min\{s.d., iqr/1.34\}$, where iqr stands for interquartile range. For both countries and all years, $s.d.$ was lower than $iqr/1.34$.

[20]For the U.K., I first used Silverman (1986) optimal bandwidth. However, the resulting income density appeared to need further smoothing at the upper tail. To correct for this, I increased the bandwidth sequentially until the income density looked smooth. This explains the use of $h = 2 \times s.d. \times n^{-1/5}$.

3.2 Estimation of Marginal Income Tax Schedules

For both countries, the marginal income tax that I use is the statutory one. I use them as a proxy of the respective marginal labor tax schedule. For the U.S., I collect the tax rates corresponding to different income brackets as published by the IRS in the "Tax Rate Schedule" from 1995 to 2004. I use the brackets corresponding to single people. To come up with a single marginal tax rate schedule for several years, I express all tax rate schedules in 2004 dollars and for every y, I take a simple average over the ten years I collected data. To maintain the constructed tax rate as a step function, I define the boundaries of the brackets as the average boundary of yearly brackets. For the U.K., I employ exactly the same procedure. This time, I collect statutory income tax rates from the "Survey of Personal Incomes Public Use Tape Documentation: Annex D: Rates of Income Tax: 1990-91 to 2004-05" located in HM-Revenue-&-Customs (1998-2007). Figure 2 shows the estimated marginal income tax schedules.

4 Results

In this section, I present my estimates of agents' jealousy parameter, α, and their contribution to the consumption externality, the ratio $\frac{\psi(\cdot)}{f(\cdot)}$.[21] This is done for the case in which the benevolent planner is utilitarian, i.e., $g(\theta) = f(\theta) \forall \theta \in \Theta$. In this instance, all taxation is Pigouvian or corrective and the jealousy parameter attains its upper bound as shown in Proposition 4. That is, $\alpha = \bar{\alpha}$. Under the assumption that the American and British societies redistribute, the jealousy parameter, α, associated with the "actual" planner's weighting density will not be higher than $\bar{\alpha}$. The estimated parameters are surprisingly moderate. For the American society, I estimate $\bar{\alpha}^{us} = 0.135$ while for the British one, I obtain $\bar{\alpha}^{uk} = 0.14$. Thus, under the

[21] The estimation uses a continuous and differentiable version of the marginal tax schedule. See Appendix D for details.

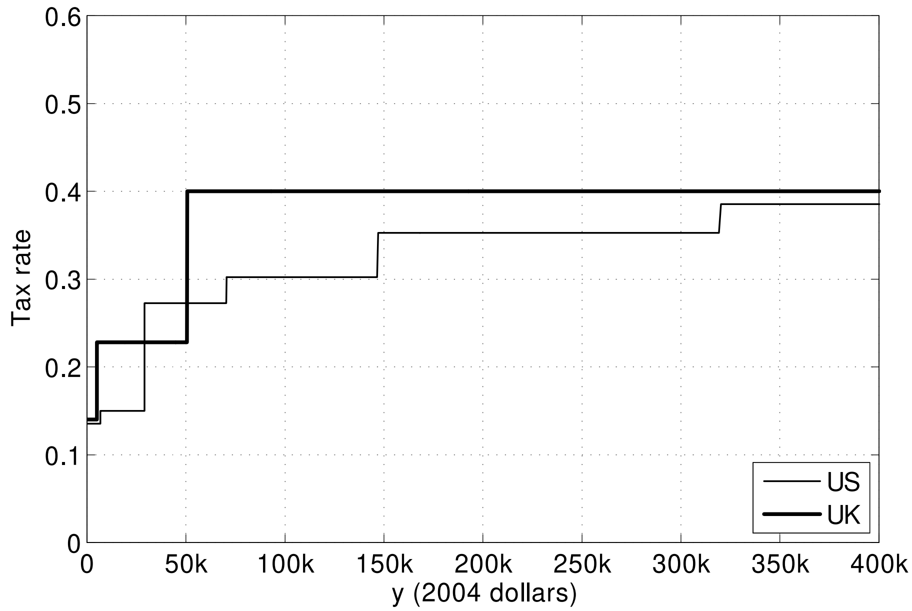

FIGURE 2: Statutory Marginal Income Tax in the U.S. and the U.K. 1995-2004

assumption that the respective planner is utilitarian, both societies seems to have very similar positional concerns.

A more intuitive interpretation of the above numbers is the following: an individual in the U.S. (U.K.) is at *least* as well off between consuming what she can purchase with one dollar (pound) than seeing the consumption of others fall by what they can buy with 13.5 cents (14 pence). Alternatively, if all agents in the U.S. (U.K.) were given one unit of the consumption good, from the point of view of a given agent such a consumption would taste at *least* like 0.865 (0.86) units after realizing that not only her but all agents increased their consumption.

Now, the question is, what is the consumption externality contribution by income in these economies? To answer this, I plot $\frac{\psi(\cdot)}{f(\cdot)}$ against the gross income cdf, $F_Y(y)$. I present my estimations under the assumption that the planner is utilitarian and that $\phi = 3$ for both countries.[22] That is, the elasticity of labor supply is $\frac{1}{3}$. This choice is in line with the work of Diamond (1998) who chooses $\phi = \{2, 5\}$ for a model with no income effects and constant elasticity of labor supply.[23]

According to Figure 3, the ratio $\frac{\psi(y)}{f(y)}$ is, roughly speaking, increasing in income for both countries. In other words, the contribution to the consumption externality is higher the more affluent individuals are.[24] More precisely, the ratio $\frac{\psi(y)}{f(y)}$ in the United Kingdom is almost flat and close to 1.75 from the third decile to the 85th percentile of the gross earnings distribution. From there on, it increases sharply reaching a level close to 4. For the United States, this ratio is close to one up to the 5th decile of the gross earnings distribution and then increases sharply reaching a level of around 2.5. This variable exhibits another sharp increase at the very upper tail of the gross

[22] Evers, Mooij, and Vuuren (2005) find that differences in estimates of labor supply elasticities across countries appear to be small. Both, U.S. and U.K. are included in their sample of countries.

[23] This choice is based on the work of Pencavel (1986). The results are robust qualitatively and quantitatively to elasticities within this range.

[24] This result is in line with one of the findings of Blanchflower and Oswald (2004).

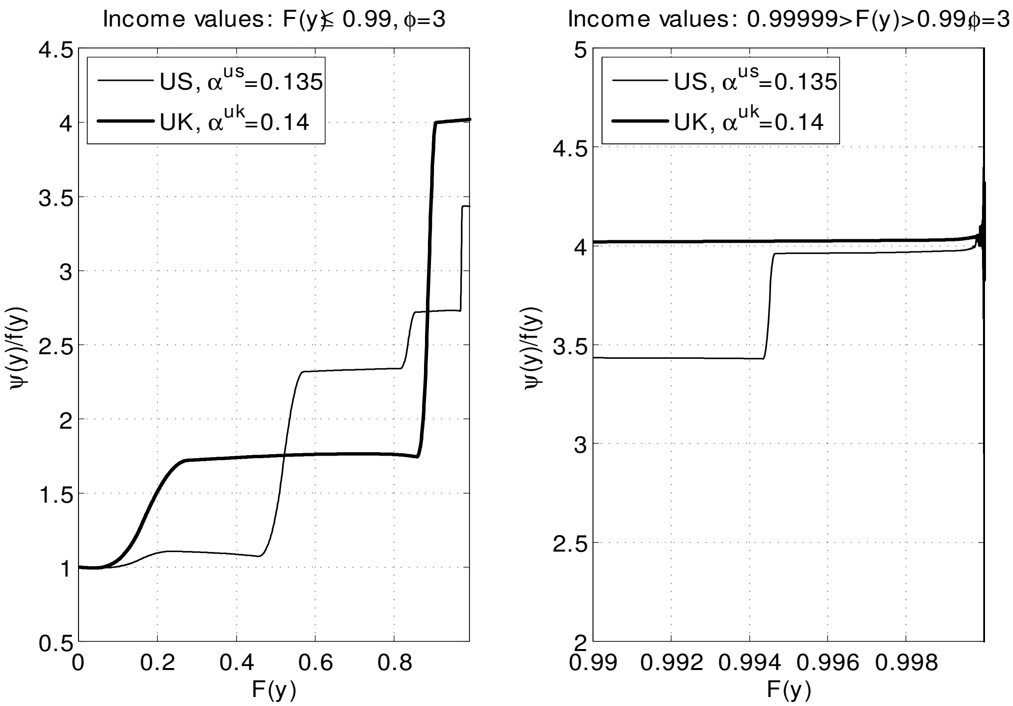

FIGURE 3: Contribution to the Consumption Externality in the U.S. and the U.K. 1995-2004

distribution where it hits a level close to 4. Not surprisingly, the ratio $\frac{\psi(y)}{f(y)}$ at the top of the earnings distribution is higher than one as stated in Proposition 5. These estimations suggest that the average consumption may not be an accurate consumption reference point but rather a weigthed average where more affluent individuals are weighted higher.

5 Conclusions

In this article I have presented a model that rationalizes high labor income taxes on affluent individuals: taxation at the high end of the labor earnings distribution may occur due to corrective considerations. This happens in the absence of a non-linear consumption tax schedule. Surprisingly, the estimated parameters that capture what is known in the literature as "jealousy" for the U.S. and the U.K. are moderate, yet producing quantitatively high effects over labor income tax rates.

Rationalizing observed labor income taxes as Pigouvian requires that the consumption externalities exerted by individuals be increasing in income. In other words, in the light of this model, observed income taxes in the U.S. and the U.K. are optimal if more affluent individuals generate a higher consumption externality than individuals with lower income and the government corrects this externality.

In future work it is important to analyze, at least numerically, how robust the results of this paper are once the quasilinearity assumption is abandoned. It would also be interesting to explore to what extent the presumed higher contribution of affluent consumers to the consumption externality is a result of these agents having access to consumption goods with higher *positional* effects and the government not being able to tax these goods directly.[25] This line of research is currently being explored in Samano (2009). Further re-

[25] In Frank (2008)'s terminology, a good is positional if its valuation depends highly on the context.

search is also necessary to understand whether positional concerns are purely instrumental as in Cole, Mailath, and Postlewaite (1992) and Postlewaite (1998) or "hard-wired" in human beings. The latter is the hypothesis of Maccheroni, Marinacci, and Rustichini (2009). Finally, as Hopkins (2008) points out, additional empirical research is needed to examine how the effect of others' income varies across the income distribution. In this paper, the parameter that captures this effect has been kept constant across agents.

References

Abel, A. (1990): "Asset prices under habit formation and catching up with the Joneses," *The American Economic Review*, 80(2), 38–42.

——— (2005): "Optimal Taxation when Consumers Have Endogenous Benchmark Levels of Consumption," *Review of Economic Studies*, 72(1), 21–42.

Alpizar, F., F. Carlsson, and O. Johansson-Stenman (2005): "How much do we care about absolute versus relative income and consumption?," *Journal of Economic Behavior and Organization*, 56(3), 405–421.

Atkinson, A. (1990): "Public economics and the economic public," *European Economic Review*, 34(2-3), 225–248.

Bault, N., G. Coricelli, and A. Rustichini (2008): "Interdependent utilities: How social ranking affects choice behavior," *PLoS ONE*, 3(10).

Blanchflower, D., and A. Oswald (2004): "Well-being over time in Britain and the USA," *Journal of Public Economics*, 88(7-8), 1359–1386.

Boskin, M., and E. Sheshinski (1978): "Optimal redistributive taxation when individual welfare depends upon relative income," *The Quarterly Journal of Economics*, 94(2), 589–601.

CAMPBELL, J., AND J. COCHRANE (1999): "By force of habit: A consumption-based explanation of aggregate stock market behavior," *Journal of Political Economy*, 107(2), 205–251.

CARLSSON, F., O. JOHANSSON-STENMAN, AND P. MARTINSSON (2007): "Do You Enjoy Having More than Others? Survey Evidence of Positional Goods," *Economica*, 74(4).

CLARK, A., AND A. OSWALD (1996): "Satisfaction and comparison income," *Journal of Public Economics*, 61(3), 359–381.

CLARK, A., AND C. SENIK (2008): "Who compares to whom? The anatomy of income comparisons in Europe," Working Paper 65, Paris School of Economics.

CODDINGTON, E. (1989): *An Introduction to Ordinary Differential Equations*. Courier Dover Publications.

COLE, H., G. MAILATH, AND A. POSTLEWAITE (1992): "Social Norms, Savings Behavior, and Growth," *Journal of Political Economy*, 100(6), 1092.

CONSTANTINIDES, G. (1990): "Habit Formation: A Resolution of the Equity Premium Puzzle," *Journal of Political Economy*, 98(3), 519.

DIAMOND, P. (1998): "Optimal Income Taxation: An Example with a U-Shaped Pattern of Optimal Marginal Tax Rates," *American Economic Review*, 88(1), 83–95.

DUESENBERRY, J. (1949): *Income, Saving, and the Theory of Consumer Behavior*. Harvard University Press.

DUPOR, B., AND W. LIU (2003): "Jealousy and Equilibrium Overconsumption," *American Economic Review*, 93(1), 423–428.

Dynan, K., and E. Ravina (2007): "Increasing income inequality, external habits, and self-reported happiness," *American Economic Review*, 97(2), 226–231.

Evers, M., R. Mooij, and D. Vuuren (2005): "What Explains the Variation in Estimates of Labour Supply Elasticities?," *CESifo Working Paper 1633*.

Ferrer-i Carbonell, A. (2005): "Income and well-being: an empirical analysis of the comparison income effect," *Journal of Public Economics*, 89(5-6), 997–1019.

Frank, R. (2008): "Should public policy respond to positional externalities?," *Journal of Public Economics*, 92(8), 1777–1786.

Gali, J. (1994): "Keeping up with the Joneses: Consumption Externalities, Portfolio Choice, and Asset Prices," *Journal of Money, Credit and Banking*, 26(1), 1–8.

Heaton, J. (1995): "An empirical investigation of asset pricing with temporally dependent preference specifications," *Econometrica*, 63(3), 681–717.

HM-Revenue-&-Customs (1998-2007): *Survey of Personal Incomes, Several Years: Public Use Tape [computer file]*. Colchester, Essex: UK Data Archive [distributor].

Hopkins, E. (2008): "Inequality, happiness and relative concerns: What actually is their relationship?," *Journal of Economic Inequality*, 6(4), 351–372.

Internal-Revenue-Service (1995-2004): *Statistics of Income: Individual Tax Returns*. Washington, D.C: U.S. Government Printing Office.

Ireland, N. (2001): "Optimal income tax in the presence of status effects," *Journal of Public Economics*, 81(2), 193–212.

J. Solnick, S., and D. Hemenway (1998): “Is more always better?: A survey on positional concerns,” *Journal of Economic Behavior and Organization*, 37(3), 373–383.

Ljungqvist, L., and H. Uhlig (2000): “Tax Policy and Aggregate Demand Management Under Catching Up with the Joneses,” *American Economic Review*, 90(3), 356–366.

Luttmer, E. (2005): “Neighbors as Negatives: Relative Earnings and Well-Being,” *The Quarterly Journal of Economics*, 120(3), 963–1002.

Maccheroni, F., M. Marinacci, and A. Rustichini (2009): “Social Decision Theory: Choosing within and between Groups,” Mimeo, Universita Bocconi, Collegio Carlo Alberto and University of Minnesota.

McBride, M. (2001): “Relative-income effects on subjective well-being in the cross-section,” *Journal of Economic Behavior and Organization*, 45(3), 251–278.

Mirrlees, J. (1971): “An Exploration in the Theory of Optimum Income Taxation,” *Review of Economic Studies*, 38(114), 175–208.

Neumark, D., and A. Postlewaite (1998): “Relative income concerns and the rise in married women’s employment,” *Journal of Public Economics*, 70(1), 157–183.

Oswald, A. (1983): “Altruism, jealousy and the theory of optimal non-linear taxation,” *Journal of Public Economics*, 20(1), 77–87.

Pencavel, J. (1986): “Labor Supply of Men: A Survey,” *Handbook of Labor Economics*, 1, 3–102.

Postlewaite, A. (1998): “The social basis of interdependent preferences,” *European Economic Review*, 42(3-5), 779–800.

Rustichini, A., and A. Vostroknutov (2007): "Competition with Skill and Luck," Mimeo, University of Minnesota.

Sadka, E. (1976): "On income distribution, incentive effects and optimal income taxation," *Review of Economic Studies*, 43(2), 261–8.

Saez, E. (2001): "Using Elasticities to Derive Optimal Income Tax Rates," *Review of Economic Studies*, 68(1), 205–229.

Salanié, B. (1997): *The Economics of Contracts: A Primer*. MIT Press.

Samano, D. (2009): "Optimal Linear Taxation of Positional Goods," Mimeo, University of Minnesota.

Seade, J. (1977): "On the Shape of Optimal Tax Schedules," *Journal of Public Economics*, 7(2), 203–235.

Senik, C. (2008): "Ambition and Jealousy: Income Interactions in the Old Europe Versus the New Europe and the United States," *Economica*, 75(299), 495–513.

Silverman, B. (1986): *Density Estimation for Statistics and Data Analysis*. Chapman & Hall/CRC.

Stokey, N. (2008): *Economics of Inaction: Stochastic Control Models with Fixed Costs*. Princeton University Press.

Tuomala, M. (1990): *Optimal Income Tax and Redistribution*. Oxford University Press, USA.

Veblen, T. (1899): *The Theory of the Leisure Class, 1934 edition*.

Wand, M., J. Marron, and D. Ruppert (1991): "Transformations in density estimation," *Journal of the American Statistical Association*, 86(414), 343–353.

Werning, I. (2007): "Pareto Efficient Income Taxation," Mimeo, MIT.

Appendix

A Proofs

Proof of Proposition 1

Proof. The first step is to transform the continuum of incentive compatibility constraints (4) into a first order condition. Let

$$W(\theta,\theta') \equiv u(c(\theta'),C) - v\left(\frac{y(\theta')}{\theta}\right) \tag{21}$$

A necessary condition for truthful revelation of type is $\frac{\partial W(\theta,\theta')}{\partial\theta'}|_{\theta'=\theta} = 0$, therefore it follows that

$$u_c(c(\theta),C)c'(\theta) = v'\left(\frac{y(\theta)}{\theta}\right)\frac{y'(\theta)}{\theta} \quad \forall\theta\in\Theta \tag{22}$$

Moreover, under truthful revelation $W(\theta) = u(c(\theta),C) - v\left(\frac{y(\theta)}{\theta}\right)$ and hence, $W'(\theta) = u_c(c(\theta),C)c'(\theta) - v'\left(\frac{y(\theta)}{\theta}\right)\frac{y'(\theta)}{\theta} + v'\left(\frac{y(\theta)}{\theta}\right)\frac{y(\theta)}{\theta^2}$, which together with (22) becomes

$$W'(\theta) = v'\left(\frac{y(\theta)}{\theta}\right)\frac{y(\theta)}{\theta^2} \quad \forall\theta\in\Theta. \tag{23}$$

Following Werning (2007), I define the expenditure function $e(W(\theta),y(\theta),C;\theta)$ to satisfy $W(\theta) = u(e,C) - v\left(\frac{y(\theta)}{\theta}\right)$. The planner problem can be reestated as

$$\max_{W(\cdot),y(\cdot),C}\int_\Theta W(\theta)g(\theta)d\theta \tag{24}$$

s.t

$$\int_\Theta e(W(\theta),y(\theta),C;\theta)f(\theta)d\theta = \int_\Theta y(\theta)f(\theta)d\theta \tag{25}$$

$$W'(\theta) = v'\left(\frac{y(\theta)}{\theta}\right)\frac{y(\theta)}{\theta^2} \quad \forall \theta \in \Theta \tag{26}$$

$$C = \int_\Theta e(W(\theta), y(\theta), C; \theta)\psi(\theta)d\theta \tag{27}$$

The corresponding Lagrangian is

$$\mathfrak{L}(W(\theta), y(\theta), C, \lambda, \mu(\theta), \gamma) = \int_\Theta W(\theta)g(\theta)d\theta - \lambda \int_\Theta \left[\left(e(W(\theta), y(\theta), C; \theta) - y(\theta)\right) f(\theta)\right] d\theta$$

$$+ \int_\Theta \mu(\theta)\left[W'(\theta) - v'\left(\frac{y(\theta)}{\theta}\right)\frac{y(\theta)}{\theta^2}\right] d\theta + \gamma\left[C - \int_\Theta e(W(\theta), y(\theta), C; \theta)\psi(\theta)d\theta\right] \tag{28}$$

Using integration by parts, it follows that

$$\int_\Theta \mu(\theta)W'(\theta)d\theta = \mu(\bar{\theta})W(\bar{\theta}) - \mu(\underline{\theta})W(\underline{\theta}) - \int_\Theta \mu'(\theta)W(\theta)d\theta \tag{29}$$

thus, we can reexpress the above Lagrangian as

$$\mathfrak{L}(W(\theta), y(\theta), C, \lambda, \mu(\theta), \gamma) = \int_\Theta W(\theta)g(\theta)d\theta - \lambda \int_\Theta \big[(e(W(\theta), y(\theta), C; \theta) - y(\theta))f(\theta)\big] d\theta$$

$$+\mu(\bar{\theta})W(\bar{\theta}) - \mu(\underline{\theta})W(\underline{\theta}) - \int_\Theta \mu'(\theta)W(\theta)d\theta - \int_\Theta \mu(\theta)v'\left(\frac{y(\theta)}{\theta}\right)\frac{y(\theta)}{\theta^2}d\theta$$

$$+\gamma\left[C - \int_\Theta e(W(\theta), y(\theta), C; \theta)\psi(\theta)d\theta\right] \tag{30}$$

Assuming interior solution, it follows from first order conditions that

$W(\theta)$:

$$g(\theta) - \lambda f(\theta) e_W(W(\theta), y(\theta), C; \theta) - \mu'(\theta) - \gamma\psi(\theta) e_W(W(\theta), y(\theta), C; \theta) = 0 \tag{31}$$

$y(\theta)$:

$$-\lambda e_y(W(\theta), y(\theta), C; \theta) f(\theta) + \lambda f(\theta) - \frac{\mu(\theta)}{\theta^2} v'\left(\frac{y(\theta)}{\theta}\right)\left[1 + \frac{1}{\epsilon(\theta)}\right]$$

$$-\gamma e_y(W(\theta), y(\theta), C; \theta)\psi(\theta) = 0 \tag{32}$$

C :

$$-\lambda \int_\Theta e_C(W(\theta), y(\theta), C; \theta) f(\theta) d\theta + \gamma - \gamma \int_\Theta e_C(W(\theta), y(\theta), C; \theta)\psi(\theta) d\theta = 0 \tag{33}$$

together with the boundary conditions $\mu(\underline{\theta}) = \mu(\bar{\theta}) = 0$ and where $\epsilon(\theta) \equiv \frac{v'(\frac{y(\theta)}{\theta})}{v''(\frac{y(\theta)}{\theta})\frac{y(\theta)}{\theta}}$. Moreover, implicitly differentiating $W(\theta) = u(e, C) - v\left(\frac{y(\theta)}{\theta}\right)$ we have that $e_W(W(\theta), y(\theta), C; \theta) = \frac{1}{u_c(c(\theta), C)}$, $e_y(W(\theta), y(\theta), C; \theta) = \frac{v'\left(\frac{y(\theta)}{\theta}\right)\frac{1}{\theta}}{u_c(c(\theta), C)}$ and $e_C(W(\theta), y(\theta), C; \theta) = -\frac{u_C(c(\theta), C)}{u_c(c(\theta), C)}$. The result follows after manipulating (31)-(33). □

Proof of Proposition 2

Proof. Taking first order conditions in agent's problem we have

$$\frac{T'(y(\theta))}{1 - T'(y(\theta))} = \frac{u_c(c^{eq}(\theta), C^{eq})}{v'\left(\frac{y^{eq}(\theta)}{\theta}\right)\frac{1}{\theta}} - 1 \tag{34}$$

where $C^{eq} = \int_\Theta c^{eq}(\theta)\psi(\theta)d\theta$. Substituting (11) into (34) it follows that

$$\frac{u_c(c^{eq}(\theta)), C^{eq})}{v'(\frac{y^{eq}(\theta)}{\theta})\frac{1}{\theta}} - 1 = \frac{\gamma\psi(\theta)}{\lambda f(\theta)} +$$

$$\frac{u_c(c^*(\theta), C^*)}{\theta f(\theta)} \left[1 + \frac{1}{\epsilon^*(\theta)}\right] \int_{\underline{\theta}}^{\theta} \left[\frac{g(t)}{\lambda} - \frac{f(t)}{u_c(c^*(t), C^*)} - \frac{\gamma}{\lambda} \frac{\psi(t)}{u_c(c^*(t), C^*)}\right] dt \tag{35}$$

Since in equilibrium the government balances its budget, we must have that

$$\int_{\Theta} c^{eq}(\theta) f(\theta) d\theta = \int_{\Theta} y^{eq}(\theta) f(\theta) d\theta \tag{36}$$

thus from (35)-(36) we conclude that $\{c^{eq}(\theta), y^{eq}(\theta)\}_{\theta \in \Theta} = \{c^*(\theta), y^*(\theta)\}_{\theta \in \Theta}$.

□

Proof of Proposition 3

Proof. By quasi-linearity of preferences $u_c(c(\theta), C) = 1$. Thus, using expression (11), the optimal income tax satisfies

$$\frac{T'(y^*(\theta))}{1 - T'(y^*(\theta))} = \frac{\gamma}{\lambda} \frac{\psi(\theta)}{f(\theta)} + \frac{1}{\theta f(\theta)} \left[1 + \frac{1}{\epsilon^*(\theta)}\right] \frac{\mu(\theta)}{\lambda} \tag{37}$$

with

$$\mu(\theta) = \int_{\underline{\theta}}^{\theta} [g(t) - \lambda f(t) - \gamma \psi(t)] dt. \tag{38}$$

By boundary conditions we have $\mu(\bar{\theta}) = 0$, hence

$$\int_{\Theta} [g(t) - \lambda f(t) - \gamma \psi(t)] dt = 0 \quad \Rightarrow \quad \lambda = 1 - \gamma \tag{39}$$

which together with the fact that $\frac{\gamma}{\lambda} = \frac{\alpha}{1-\alpha}$ which follows from expression (7) when preferences are quasi-linear implies that $\lambda = 1 - \alpha$ and $\gamma = \alpha$. Plugging the previous values into (37) and substituting into (38) delivers the result after algebraic manipulations. □

B Asymptotic Tax in a Quasi-Linear Environment

Under the assumption that $\bar{\theta} < \infty$ it can be seen from Corollary 1 that $\frac{T'(y^*(\bar{\theta}))}{1-T'(y^*(\bar{\theta}))} = \frac{\alpha}{(1-\alpha)} \frac{\psi(\bar{\theta})}{f(\bar{\theta})}$. Hence, under a bounded distribution of skills I obtain a non-zero taxation at the top due to corrective considerations. Proposition 6 exhibits the formula for the optimal marginal labor income tax as $\bar{\theta}$ goes to infinity. I consider the case of quasi-linear preferences, a constant elasticity of labor supply and $f(\theta)$ Pareto-distributed. The last assumption is used based on Diamond (1998) and Saez (2001). Both studies obtain positive asymptotic marginal tax rates, however these results depend critically on $f(\theta)$ being Pareto.[26]

Proposition 6. *Suppose $f(\theta)$ is Pareto distributed with parameter $k > 0$ and that $L_1 \equiv \lim_{\theta\to\infty} \frac{\psi(\theta)}{f(\theta)}$ and $L_2 \equiv \lim_{\theta\to\infty} \frac{g(\theta)}{f(\theta)}$ exist. If $u(c(\theta), C) = c(\theta) - \alpha C$, $\alpha \in [0, 1)$, $v\left(\frac{y}{\theta}\right) = \frac{1}{1+\phi}\left(\frac{y}{\theta}\right)^{1+\phi}, \phi > 0$ then $\frac{T'_\infty}{1-T'_\infty} = \frac{\alpha}{(1-\alpha)} L_1 \left[\frac{1+\phi+k}{k}\right] + \frac{1+\phi}{k}\left[1 - \frac{1}{(1-\alpha)} L_2\right]$ where $T'_\infty \equiv \lim_{\theta\to\infty} T'(y^*(\theta))$.*

Proof. From Proposition (3) and simple algebraic manipulations we have

$$\frac{T'(y^*(\theta))}{1 - T'(y^*(\theta))} = \frac{\alpha}{(1-\alpha)} \frac{\psi(\theta)}{f(\theta)} + \frac{1 - F(\theta)}{(1-\alpha)\theta f(\theta)} \left[1 + \frac{1}{\epsilon^*(\theta)}\right] \times$$

$$\frac{[G(\theta) - (1-\alpha)F(\theta) - \alpha\Psi(\theta)]}{1 - F(\theta)} \tag{40}$$

Since $f(\theta)$ is Pareto it follows that $\frac{1-F(\theta)}{\theta f(\theta)} = \frac{1}{k}$ and since $v(\frac{y}{\theta}) = \frac{1}{1+\phi}\left(\frac{y}{\theta}\right)^{1+\phi}$ we have that $\left[1 + \frac{1}{\epsilon^*(\theta)}\right] = 1 + \phi$. Using the fact that $L_1 \equiv \lim_{\theta\to\infty} \frac{\psi(\theta)}{f(\theta)} < \infty$

[26] For a Pareto distribution $f(\theta) = \frac{k\underline{\theta}^k}{\theta^{k+1}}, \theta \in [\underline{\theta}, \infty), \underline{\theta} > 0, k > 0$ and $F(\theta) = 1 - \left(\frac{\underline{\theta}}{\theta}\right)^k$.

it follows that

$$\lim_{\theta\to\infty}\frac{T'(y^*(\theta))}{1-T'(y^*(\theta))}=\frac{\alpha}{1-\alpha}L_1+\frac{(1+\phi)}{k(1-\alpha)}\lim_{\theta\to\infty}\frac{\left[G(\theta)-(1-\alpha)F(\theta)-\alpha\Psi(\theta)\right]}{1-F(\theta)} \tag{41}$$

Using L'Hôpital's rule, $\lim_{\theta\to\infty}\frac{\left[G(\theta)-(1-\alpha)F(\theta)-\alpha\Psi(\theta)\right]}{1-F(\theta)}=-L_2+(1-\alpha)+\alpha L_1$ and substituting into (41) delivers the result. □

Corollary 2 (Proposition 3)**.** *If $L_1\geq 1$ and $L_2\leq 1$, then $T'_\infty\geq\alpha$.*

Proof. Trivial. □

Figure 4 shows the ratio $\frac{1-F_Y(y)}{yf_Y(y)}$ of the income distribution in the U.S. for 1992, 1993 and from 1995-2004. I include the years 1992 and 1993 since Saez (2001) estimates a Pareto distribution parameter for labor earnings based on these years. This ratio was constructed after smoothing the upper tail of $f_Y(y)$ using a gaussian kernel. The width was set at $h=1.36\times s.d.\times n^{-1/5}$, where the standard deviation ($s.d.$) and n were calculated for observations exceeding 13.5 million dollars (expressed in 2004 dollars). Table 1 reports the number of observations, n, in the sample exceeding that threshold.[27]

Table 1: In Sample Number of Gross Income Observations Exceeding \$13,500,000 (2004 dollars)

Year	Observations	Year	Observations
1992	153	1999	486
1993	75	2000	666
1995	78	2001	405
1996	134	2002	261
1997	217	2003	282
1998	333	2004	391

It can be seen that at least for some years the ratio $\frac{1-F_Y(y)}{yf_Y(y)}$ at the very top of the distribution is decreasing. This fact indicates that the very top of the

[27] This amount is approximately equivalent to 10 million expressed in 1992 dollars. Saez (2001) reports that starting at this income level the number of taxpayers in the database is very small.

income distribution of the United States may not be accurately represented by a Pareto distribution. Moreover, a decreasing $\frac{1-F_Y(y)}{yf_Y(y)}$ would imply a decreasing pattern of optimal taxes in the canonical Mirrleesian model. This is indeed what Mirrlees (1971) finds since he assumes a log-normal distribution of skills for which the ratio is decreasing. The model with consumption externalities would deliver asymptotic non-zero optimal taxes even if the ratio $\frac{1-F_Y(y)}{yf_Y(y)}$ is decreasing.

C Proof of Proposition 4

Proof. Observe that we can reexpress (16) as[28]

$$\alpha = \frac{\int_\Theta \frac{f(\theta)T'(y(\theta))}{\theta^{1+\phi}(1-T'(y(\theta)))}d\theta + (1+\phi)\int_\Theta \frac{(F(\theta)-G(\theta))}{\theta^{2+\phi}}d\theta}{\int_\Theta \frac{f(\theta)T'(y(\theta))}{\theta^{1+\phi}(1-T'(y(\theta)))}d\theta + (1+\phi)\int_\Theta \frac{F(\theta)}{\theta^{2+\phi}}d\theta + \frac{1}{\bar{\theta}^{1+\phi}}} \tag{45}$$

Thus,

$$\bar{\alpha} = \max_{G(\cdot)}\{\alpha \mid G(\theta) \geq F(\theta)\ \forall\theta \in \Theta, G'(\theta) \geq 0, G(\underline{\theta}) = 0, G(\bar{\theta}) = 1\}$$

---

[28] To see this, observe that

$$\int_\Theta \frac{f(\theta)}{\theta^{1+\phi}(1-T'(y(\theta)))}d\theta = \int_\Theta \frac{f(\theta)T'(y(\theta))}{\theta^{1+\phi}(1-T'(y(\theta)))}d\theta + \int_\Theta \frac{f(\theta)}{\theta^{1+\phi}}d\theta \tag{42}$$

Using integration by parts it follows that

$$(1+\phi)\int_\Theta \frac{F(\theta)}{\theta^{\phi+2}}d\theta = -\frac{1}{\bar{\theta}^{1+\phi}} + \int_\Theta \frac{f(\theta)}{\theta^{1+\phi}}d\theta \tag{43}$$

Plugging (43) into (42) we obtain that the denominator of (45) is equal to the denominator of (16). To see that the numerator of (45) is equal to the one in (16) use expression (42) together with the fact that by integration by parts

$$\int_\Theta \frac{g(\theta)}{\theta^{1+\phi}}d\theta = (1+\phi)\int_\Theta \frac{G(\theta)}{\theta^{\phi+2}}d\theta + \frac{1}{\bar{\theta}^{1+\phi}} \tag{44}$$

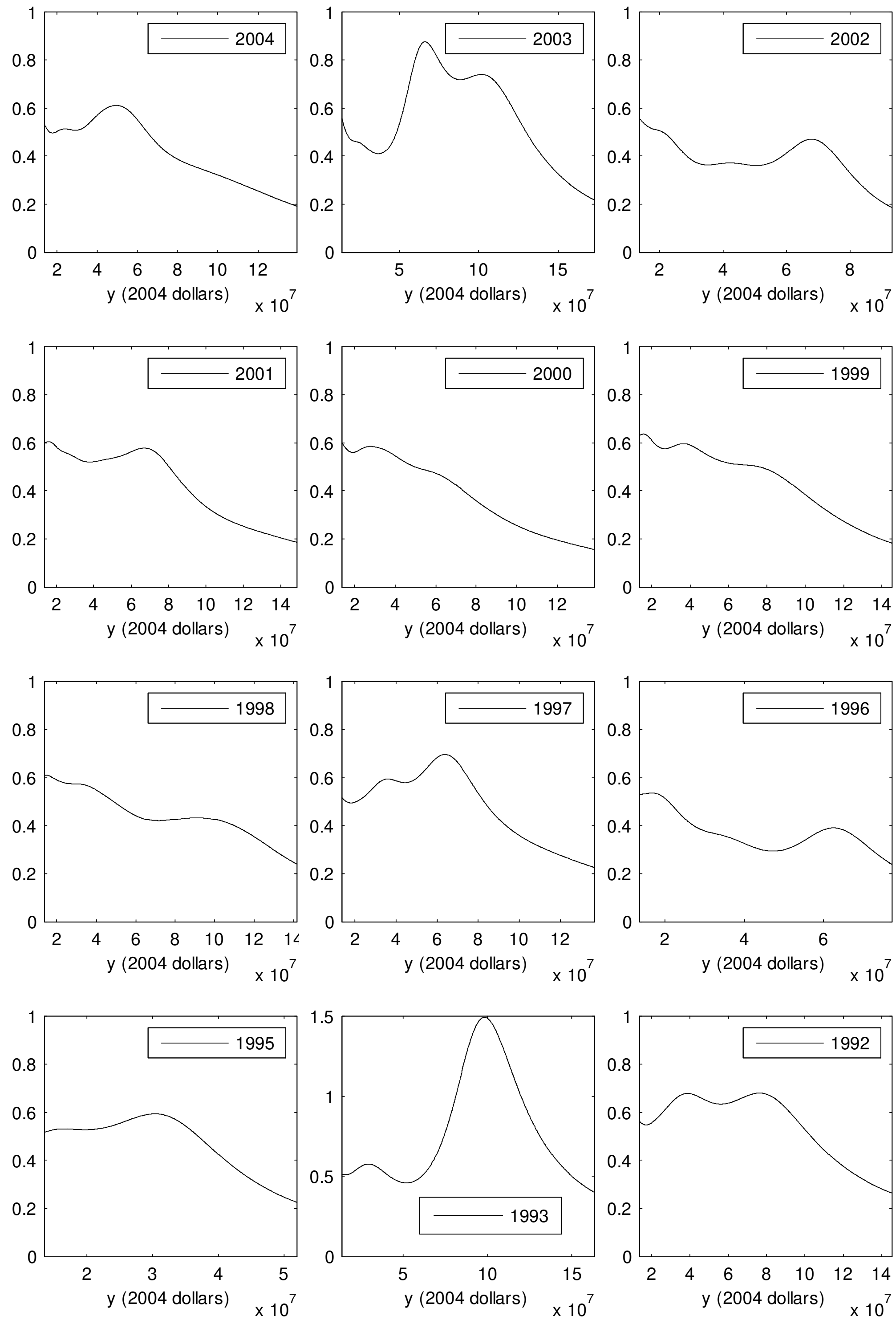

FIGURE 4: Kernel smoothed $\frac{1-F_Y(y)}{yf_Y(y)}$ ratio in the U.S. $y \geq \$13,500,000$ (2004 dollars)

To obtain $G(\cdot)$, I solve the following maximization problem

$$\max_{G(\cdot)} \int_{\Theta} \frac{(F(\theta) - G(\theta))}{\theta^{2+\phi}} d\theta$$

s.t.

$$G(\theta) \geq F(\theta) \forall \theta \in \Theta \tag{46}$$

$$G'(\theta) \geq 0, \quad G(\underline{\theta}) = 0, \quad G(\bar{\theta}) = 1 \tag{47}$$

Let me consider the optimization problem ignoring constraints (47). The corresponding Lagrangian is

$$\mathfrak{L}(G(\theta), \gamma(\theta); F(\theta), \phi) = \int_{\Theta} \frac{(F(\theta) - G(\theta))}{\theta^{2+\phi}} d\theta + \int_{\Theta} \gamma(\theta)[G(\theta) - F(\theta)] d\theta$$

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

$$\frac{1}{\theta^{2+\phi}} = \gamma(\theta) \quad \forall \theta \in \Theta \tag{48}$$

Clearly, $\gamma(\theta) > 0 \forall \theta \in \Theta$. By the slackness condition, we must have

$$\gamma(\theta)[G(\theta) - F(\theta)] = 0 \quad \forall \theta \in \Theta \tag{49}$$

thus, $G(\theta) = F(\theta) \forall \theta \in \Theta$. Clearly this solution satisfies (47). Finally, substituting $g(\theta) = f(\theta)$ into (16) we obtain

$$\bar{\alpha} = \frac{\int_{\Theta} \frac{f(\theta) T'(y(\theta))}{\theta^{1+\phi}(1-T'(y(\theta)))} d\theta}{\int_{\Theta} \frac{f(\theta)}{\theta^{1+\phi}(1-T'(y(\theta)))} d\theta}.$$

Making a change of variable using (17), it can be shown that

$$\bar{\alpha} = \frac{\int_{\Phi^{-1}(\underline{\theta})}^{\Phi^{-1}(\bar{\theta})} \frac{f_Y(y)T'(y)}{y^{\phi}} dy}{\int_{\Phi^{-1}(\underline{\theta})}^{\Phi^{-1}(\bar{\theta})} \frac{f_Y(y)}{y^{\phi}} dy} = \frac{E_Y[y^{-\phi}T'(y)]}{E_Y[y^{-\phi}]} \tag{50}$$

$\square$

Proof of Proposition 5

Proof. Since $T'(y(\bar{\theta})) \geq T'(y(\theta)) \forall \theta \in \Theta$ with strict inequality for some interval I, it follows from (50) that

$$\bar{\alpha} = \frac{E_Y[y^{-\phi}T'(y)]}{E_Y[y^{-\phi}]} < \frac{E_Y[y^{-\phi}]T'(\bar{y})}{E_Y[y^{-\phi}]} = T'(y(\bar{\theta}))$$

where the strict inequality comes from the fact that $a \geq \underline{\theta}$ and $b < \bar{\theta}$. Since, $\frac{T'(y(\bar{\theta}))}{1-T'(y(\bar{\theta}))} = \frac{\alpha}{1-\alpha}\frac{\psi(\bar{\theta})}{f(\bar{\theta})}$ it follows that $\frac{\bar{\alpha}}{1-\bar{\alpha}} < \frac{\alpha}{1-\alpha}\frac{\psi(\bar{\theta})}{f(\bar{\theta})}$, thus $\frac{\psi(\bar{\theta})}{f(\bar{\theta})} > 1$. $\square$

D Constructing a Continuous and Differentiable Marginal Tax Schedule

In this Appendix, I construct a continuous and differentiable version of the statutory marginal tax schedule that I employ for my estimation of $\psi(\theta)$. I base this procedure in an insight from Stokey (2008). Without loss of generality, I will focus on a continuous and differentiable version of a statutory tax schedule with two income brackets of the form

$$T'(y) = \begin{cases} c_1 & \text{if } y \leq \tilde{y} \\ c_2 & \text{if } y > \tilde{y} \end{cases} \tag{51}$$

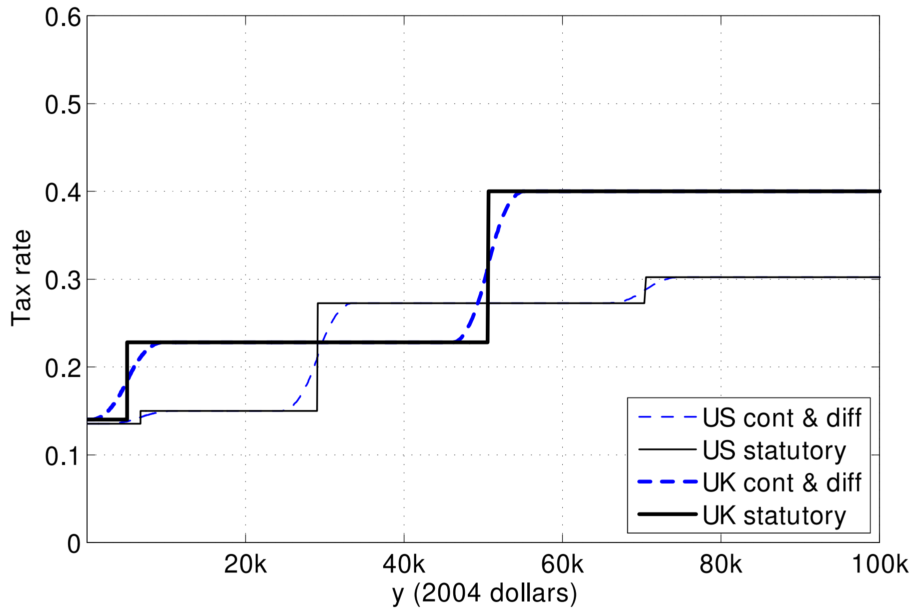

FIGURE 5: Statutory and Smoothed Marginal Income Tax in the U.S. and the U.K. 1995-2004

For $\epsilon > 0$, a continuous and differentiable version of the previous "step" tax is given by

$$T'(y) = \begin{cases} c_1 & \text{if } y \leq \tilde{y} - \frac{\epsilon}{2} \\ c_1 + \frac{2(c_2 - c_1)}{\epsilon^2}(y - \tilde{y} + \frac{\epsilon}{2})^2 & \text{if } \tilde{y} - \frac{\epsilon}{2} < y < \tilde{y} \\ c_1 + \frac{(c_2 - c_1)}{2} + \frac{2(c_2 - c_1)}{\epsilon^2}(2(\tilde{y} + \frac{\epsilon}{2})y - y^2 - \tilde{y}^2 - \tilde{y}\epsilon) & \text{if } \tilde{y} \leq y < \tilde{y} + \frac{\epsilon}{2} \\ c_2 & \text{if } y \geq \tilde{y} + \frac{\epsilon}{2} \end{cases} \tag{52}$$

Figure 5 shows the continuous and differentiable versions of the statutory taxes for the U.S. and the U.K. I use $\epsilon = 9,000$.