Income maintenance and labor force participation

Guy Laroque[1]

First draft May 2002
Revised August 2004

[1]CREST-INSEE and CNRS URA 2200. I have benefited from numerous discussions with Philippe Choné, Christian Gouriéroux, Thierry Magnac and Bernard Salanié and from the remarks of seminar participants at CREST, Université de Cergy, University of Ghent, Université d'Orléans, the Roy seminar, and at the meetings of the Econometric Society in Brisbane, Lahore, Los Angeles, Sao Paulo and Venice. Comments from Antoine d'Autume, Costas Azariadis, Richard Blundell, Eddie Dekel, Marc Fleurbaey, James Heckman, Jean-Michel Grandmont, Anne Laferrère, Rafael Repullo and three anonymous referees are gratefully acknowledged. Part of the theory stems from my joint work with Philippe Choné.

Income maintenance and labor force participation

Guy Laroque
INSEE-CREST J360, 15 boulevard Gabriel Péri, 92245 Malakoff, France

Abstract

The paper studies the optimal tax-subsidy schedules in an economy where the only decision of the agents is to work, or not, with an application to the case of France.

Given an income guarantee provided by the welfare state, the tax schedule that maximizes government revenue provides a benchmark, the Laffer bound, above which it is inefficient to tax. In fact, under mild conditions, a feasible allocation is second best optimal if and only if the associated taxes are lower than the Laffer bound. The only restriction that efficiency puts on the shape of the tax scheme is this upper Laffer bound.

The Laffer tax scheme itself can be computed from the joint distribution of the agents' productivities and work opportunity costs. Depending on the economy, it can take widely different forms, and exhibit, for instance, *negative* marginal tax rates.

After estimating the joint distribution of productivities and work opportunity costs on French data, I compute the Laffer bound for two sub-populations, single women and married women with two children or more. Quite surprisingly, the actual incentives to work appear to be very close to the bound.

Keywords: extensive margin, optimal taxation, incentives.

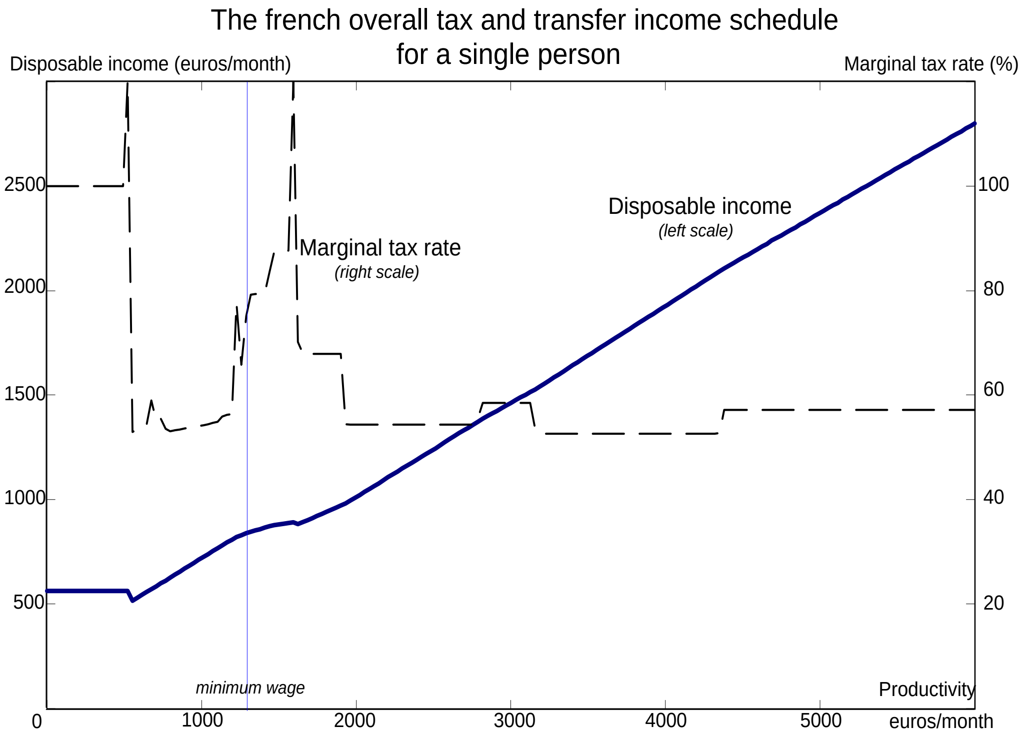

Figure 1: The French tax and transfer schedule in 1999

1 Introduction

The welfare state is often blamed for pushing the low skilled population out of the work force. Figure 1, which represents the overall long run tax and transfer system in France in 1999, plots disposable income as a function of labor cost for a single person. When not working, the maintenance income (RMI), together with average housing subsidies, amount to 560 euros per month. An unskilled worker, having the opportunity to take a full time job paid at the minimum wage (cost to the employer 1300 euros per month), would earn after transfers and taxes 840 euros per month, so that her financial incentive to work is 280 euros per month, or less than 2 euros per hour. The left part of the curve is horizontal, corresponding to a 100% marginal tax rate: working half time at the minimum wage does not increase income at all. Unskilled workers may be trapped out of the labor force or induced to join the underground economy. Such poverty traps have been the subject of a lot of attention from economists and policy makers in the past thirty years around the world (see Moffitt (2002) for a recent survey, Eissa and Liebman (1996) and Blundell, Duncan, McCrae, and Meghir (2000) for studies on the US and the UK respectively). In the United States, the Earned Income Tax Credit, followed by the welfare reform of 1996, has been motivated in part by a willingness to *make work pay* and to reduce the undesirable side effects of the Welfare State. Canada has been in the forefront in the design of schemes to induce long term unemployed persons to participate full time in the labor force, see Robins and Michalopoulos (2001) or Card, Michalopoulos, and Robins (1999).

The economists have been concerned with the distortions induced by the tax system since the profession exists. In theory, optimal taxation, provided that society's preferences for redistribution are elucidated, should be a useful guide. In practice the normative approach has not been very fruitful. Indeed, the relevant framework of optimal taxation, which goes back to the seminal paper of Mirrlees (1971), seems too far from the tax-benefit systems observed in practice to be a useful guide for policy[1]. When effort depends on financial incentives, at the *intensive* margin, the standard result has a zero marginal tax rate on the rich (which goes contrary to the common idea of equity, and is not observed). The marginal tax rate is always non negative, which rules out pushing people to work through an earnings subsidy, as intended by the EITC. The practical schemes have therefore mostly been influenced, on the one hand by political pressures which in the recent years advocate a retreat of the welfare state, and on the other hand, by empirical results which show the importance of details in welfare implementation and argue in favor of targeting specific populations.

The purpose of the present paper is to make a step in bringing together the theoretical approach of social choice theory and optimal taxation with the empirical research on labor supply. On the theory side, the paper focuses on labor supply at the *extensive* margin where the agents' decision is zero-one, to work or not to work, as in the studies of Diamond (1980), Beaudry and Blackorby (1997), Saez (2002) and Choné and Laroque (2001). In a number of countries, including France, the distribution of hours worked per week is essentially concentrated on two modes, full time and half time, so that the *intensive* margin model, where work time is adjusted continuously, seems less relevant than the *extensive*. I characterize the set of second best allocations and the tax schemes that support them, as Stiglitz (1982) did for the intensive model. Given a minimum income guarantee, one can compute a revenue maximizing tax scheme, which I refer to as the Laffer bound (Canto, Joines, and Laffer (1982)). Under mild conditions, a feasible allocation is second best optimal if and only if the taxes that implement it are lower than the Laffer bound. The Laffer bound itself is determined by the joint distribution of productivities and *work opportunity costs* (or monetary *disutilities for work*) in the economy, and theory puts essentially no restrictions on its shape (Choné and Laroque (2001)).

It follows that reforms, say towards making work pay, should be analyzed as political redistribution, provided that one stays on the right side of the Laffer bound. To measure this bound, one needs to estimate the joint distribution of

[1]Also, the optimal tax program is quite difficult to solve. This difficulty has been overcome by Saez (2001) who uses the available empirical evidence on the shape of the wage distribution and labor supply elasticities to compute optimal tax schedules.

work opportunity costs and productivities in the economy. As an illustration, I rely on a labor supply model developed with Bernard Salanié on French data. It is estimated on a sample of women aged 25-49. Work opportunity costs depend on the income of the spouse (if any) and on the number and ages of the children, as well as on an unobserved heterogeneity term. I discuss the identification of the distribution of this term and estimate the model.

Once the joint distribution of productivities and work opportunity costs is recovered from the data, it is easy to apply the theoretical computations to the particular case at hand to derive the Laffer bound, keeping fixed the minimum income guarantee when not working. The results are presented for single women, and for married women with two children or more. Quite surprisingly, the actual French tax schedule, while efficient, appears to be very close to the bound. It looks as if the interactions between the multiple agencies that shape the income tax schedule in France lead to a Leviathan state that extracts the maximum possible surplus from the population.

These results call for independent confirmation. More generally, the present work raises more questions than it answers. A major step forward would be to deal with the intensive margin (part time work), incorporating some of the standard optimal taxation literature, along the lines of Saez (2001). Another useful extension would go from the static setup to a dynamic environment. It would also be interesting to repeat the empirical exercise for other countries. The relative size of the government in the economy is much smaller in the US than in France. This type of analysis makes it possible to assess whether this contrast is all due to a difference in political attitudes as often claimed, i.e. the Laffer bounds are similar but the actual tax scheme is farther from the bound in the US than in France, or whether part of it can be explained by differences in tastes for work on both sides of the Atlantic.

2 Theory

2.1 The model

We consider an economy made of a continuum of agents. A typical agent is described by a set of exogenous characteristics, denoted by $a = (w, x, y)$. The first coordinate of a, w, denotes her productivity, while the other characteristics (x, y), together with productivity, describe her tastes for leisure or non market work. When working, the typical agent produces a quantity w of an undifferentiated desirable commodity. The characteristics x of the agent are assumed to be observable by the government and verifiable. Furthermore the democratic process allows benefits and taxes to be conditioned on the values of x. For instance, x may include the number and ages of children in the household. On the other hand, the characteristics y are private, and benefits or taxes cannot be made conditional on y. In the second best environment which I shall be considering, for each individual the government only observes x and the productivity w if she works. On the other hand, the government knows the joint distribution of (w, x, y) in the overall population, and therefore the conditional distributions, given the observables, as well.

Formally, the characteristics $a = (w, x, y)$ of the agents belong to a set A in the n dimensional Euclidean space. An economy is defined by a probability measure on A, with c.d.f. F. The aggregate resources in the economy are assumed to be finite, i.e. w is integrable with respect to the measure F.

The only choice of the agent in our model is whether to participate, or not, in the work force. The participation status of agent a is described with a function $s(a)$, where $s(a)$ is equal to 0 (no work) or 1 (work). When agent $a = (w, x, y)$ participates ($s(a) = 1$), she produces w units of commodity, while she does not produce any marketable good when she does not participate ($s(a) = 0$).

The agents' behavior is described through a measure of their disutility of work, their *work opportunity cost*[2]. Income, or consumption, is assumed to always be desirable. Consider an agent a who is indifferent between working with income $\mathbf{C}$ and not working with income $\mathbf{c}$. Her work opportunity cost, in either of these two situations, is the (possibly negative) difference $\mathbf{C} - \mathbf{c}$. The labor supply of the agent is fully characterized by the value of her work opportunity cost, which can be measured alternatively as a function of income out of work, $\Delta(\mathbf{c}; a)$, or of income at work, $\Gamma(\mathbf{C}; a)$. By definition:

$$\mathbf{C} - \mathbf{c} = \Delta(\mathbf{c}; a) = \Gamma(\mathbf{C}; a),$$

so that

$$\Delta(\mathbf{C} - \Gamma(\mathbf{C}; a); a) = \Gamma(\mathbf{C}; a),$$

and

$$\Delta(\mathbf{c}; a) = \Gamma(\mathbf{c} + \Delta(\mathbf{c}; a); a).$$

Agent a, with income $\mathbf{c}$ when out of the labor force, is willing to work whenever she faces financial incentives larger than $\Delta(\mathbf{c}; a)$. On the other hand, if a works with income $\mathbf{C}$, she would like to quit when the associated income loss is smaller than $\Gamma(\mathbf{C}; a)$.

To link the notion to a more traditional concept in microeconomics, let $u(c, s, a)$ denote agent a utility when she receives income c and has work status s. Then

$$u(c, 0, a) = u(c + \Delta(c; a), 1, a) \qquad u(c - \Gamma(c; a), 0, a) = u(c, 1, a).$$

[2]An alternative terminology, used in an earlier version of the paper, is *work aversion*. Another suggestion, due to Rafael Repullo, is *minimum inducement to work*.

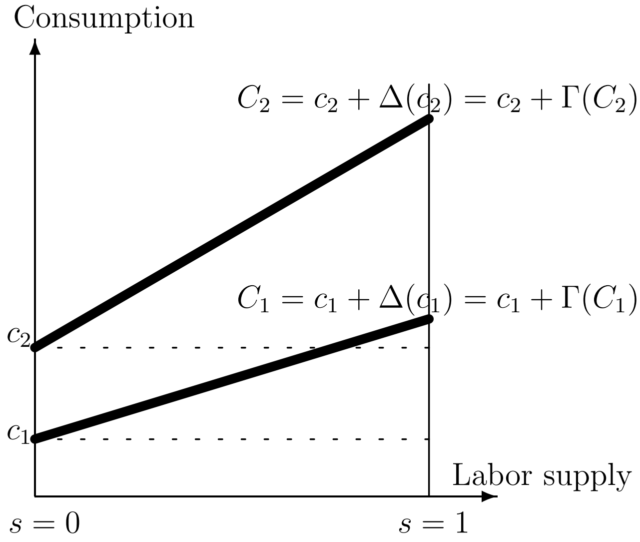

Figure 2: Indifference curves

One representation of this utility, which I shall use sometimes in the sequel of the paper, is equivalent consumption when at work:

$$u(c, s, a) = \begin{cases} c + \Delta(c; a) & \text{when } s = 0 \\ c & \text{when } s = 1. \end{cases}$$

Any monotone transformation of u would of course also be consistent with the agent's choices[3].

I assume

Assumption 1. $\Delta(c; a)$ *and* $\Gamma(c; a)$ *are defined on* $\mathbb{R}_+ \times A$ *and continuous.* $\Delta(c; a)$ *is a nondecreasing function of* c.

The larger the income when unemployed, the larger the required income *supplement* to make it worthwhile to take a job (see Figure 2). The assumption that Δ is nondecreasing in c is equivalent, in this setup, to leisure being a normal

[3]When consumption is restricted to be positive, there are some difficulties in the correspondence between the two approaches, work opportunity costs or utilities, at the lower boundary of the domain. For instance, if work opportunity cost is always strictly positive, $\Delta(0; a) > 0$, $u(c, 1, a) < u(0, 0, a)$ for all c smaller than $\Delta(0; a)$, and in this region, there is no way to make the agent indifferent between working and not working. To keep things as simple as possible, I shall assume that both $\Delta(.; a)$ and $\Gamma(.; a)$ are defined on the whole of $\mathbb{R}_+$, which could be derived from utilities defined on the real line, while restricting attention to allocations with non negative consumptions. The signs of work opportunity costs are not constrained, which accounts for agents with a negative work opportunity cost, who would rather work on the market than stay at home with the same income.

good: the supply of labor is a decreasing function of the level c of income when not working. Indeed, given a gross income at work $c+D$, the agent's labor supply is equal to zero when D is smaller than $\Delta(c;a)$, and equal to one otherwise. Then the fact that $\Delta(c;a)$ increases with c implies that labor supply decreases with c[4].

We denote by $G_{c,w,x}$ the c.d.f. of the distribution of work opportunity costs $\Delta(c;a)$ conditional on the agent productivity w and on the observable x

$$G_{c,w,x}(D) = \Pr\ (\Delta(c;a) \leq D \mid w, x).$$

Assumption 1 implies that $G_{c,w,x}(D)$ is a nonincreasing function of c.

An *allocation* describes the employment status and the income of all the agents in the economy. Formally, it is defined as a pair of integrable functions $s(a)$ and $c(a)$ with values respectively in $\{0,1\}$ and $\mathbb{R}_+$. An allocation is *individually rational* when every agent, whether working or not, is better off than in the situation where she would have a zero income: at an individually rational allocation, when an agent a works ($s(a)=1$), she is better off than not working with a zero consumption, i.e. $c(a) - \Gamma(c(a);a) \geq 0$; similarly when agent a does not work, she is better off than working with a zero consumption, $c(a) + \Delta(c(a);a) \geq 0$. An allocation $(s(.), c(.))$ is *feasible* when total consumption is equal to total production, i.e.:

$$\int c(a)\mathrm{d}F(a) = \int_{s(a)=1} w\mathrm{d}F(a). \tag{1}$$

At the *laissez-faire* allocation, the perfectly competitive wage is equal to productivity. An agent decides to work when her productivity makes it worthwhile, in comparison with a zero income when non participating, i.e. when

$$w \geq \Delta(0;a),$$

with indifference when there is equality.

Such an allocation can be very unequal, and it is of interest to look at redistribution schemes that tax the rich workers, with high w's, and give the proceeds to the unemployed. Such a redistribution scheme typically reduces the incentives to work. Indeed if $R(a)$, $R(a) \leq w$, is the after tax income of worker a, and r, $r \geq 0$, the subsistence level attributed to the unemployed, the decision to work under the redistribution scheme is associated with the inequality

$$R(a) - r \geq \Delta(r;a),$$

which is always more stringent than at the *laissez-faire* allocation. The purpose of the paper is to look at the tradeoff between equity (more equal utility levels) and efficiency (loss of output due to non participation generated by redistribution)

[4]Note that, by the definition of Γ, Assumption 1 implies that $\Gamma(c;a)$ is nondecreasing in c. Note also that $\Delta(c;a) \geq \Gamma(c;a)$ if and only if $\Delta(c,a) \geq 0$.

depending on the government objective and information and to see whether the optimal taxation schemes exhibit some general properties.

Following tradition, I study the set of optimal allocations, starting with the case of complete information of the planner (first best), following with the situation where the planner only observes part of the agents' characteristics and the allocation has to be measurable with respect to the planner's information (second best).

2.2 First best allocations

In this setup, it is easy to characterize the set of Pareto optimal allocations. Indeed, an allocation is Pareto optimal when someone works if and only if her productivity is larger than the extra income necessary to compensate her for the hardship of work.

The fact that a Pareto optimal allocation has to satisfy the above condition is easy to understand. Consider a feasible allocation $(c(a), s(a))$, and suppose that for some (group of) agent(s) $s(a) = 1$ while $\Gamma(c(a); a) > w(a)$. Then modify this allocation by putting the corresponding agents out of employment with consumption $c(a) - \Gamma(c(a); a)$, so that their utilities are unchanged. In the process, the planner saves $\Gamma(c(a), a)$ per head, but loses the production $w(a)$. From the inequality, the planner earns a positive surplus equal to $\Gamma(c(a); a) - w(a)$. Similarly, suppose that there is a group of unemployed agents ($s(a) = 0$) with $\Delta(c(a); a) < w(a)$. Putting them to work while keeping them at the same utility yields a surplus equal to $w(a) - \Delta(c(a); a)$.

Formally:

Theorem 1. *The individually rational allocation $(c(a), s(a))$ is Pareto optimal if and only if*

$$\begin{array}{l} s(a) = 1 \Longrightarrow w(a) \geq \Gamma(c(a); a) \\ s(a) = 0 \Longrightarrow w(a) \leq \Delta(c(a); a). \end{array} \tag{2}$$

Proof: It just remains to be proved that an individually rational feasible allocation $(c(a), s(a))$ which satisfies (2) is Pareto optimal. Let $(c'(a), s'(a))$ be another allocation which makes every agent at least as well off, and some strictly better off. I shall show that it is not feasible.

Consider a partition of the set of agents into four subsets, A_{ij}, $i = 0, 1$, $j = 0, 1$, where A_{ij} is the set of agents a such that $s(a) = i$ and $s'(a) = j$. In words, A_{00} is the set of agents who are unemployed in both allocations. Since consumption is desirable,

$$\int_{A_{00} \cup A_{11}} c'(a) dF(a) \geq \int_{A_{00} \cup A_{11}} c(a) dF(a),$$

and, by definition of the work opportunity cost,

$$\int_{A_{01}} c'(a)dF(a) \geq \int_{A_{01}} [c(a) + \Delta(c(a);a)]dF(a),$$

$$\int_{A_{10}} c'(a)dF(a) \geq \int_{A_{10}} [c(a) - \Gamma(c(a);a)]dF(a),$$

with some strict inequality. Summing up the three preceding inequalities yields:

$$\int c'(a)dF(a) > \int c(a)dF(a) + \int_{A_{01}} \Delta(c(a);a)dF(a) - \int_{A_{10}} \Gamma(c(a);a)dF(a).$$

From (2), it follows that:

$$\int c'(a)dF(a) > \int c(a)dF(a) + \int_{A_{01}} w(a)dF(a) - \int_{A_{10}} w(a)dF(a).$$

The initial allocation satisfies the feasibility constraint

$$\int c(a)dF(a) = \int_{A_{10}\cup A_{11}} w(a)dF(a),$$

so that

$$\int c'(a)dF(a) > \int_{A_{01}\cup A_{11}} w(a)dF(a) = \int_{s'(a)=1} w(a)dF(a),$$

which proves that the allocation $(c'(a), s'(a))$ is not feasible. ∎

At a Pareto optimal allocation, every one with a work opportunity cost larger than her productivity is kept out of the work force; all those whose work opportunity costs are smaller than their productivities work. Pareto optimality implies full efficiency of the allocation of time between market and non market activities. It follows that the *laissez-faire* equilibrium is Pareto optimal. Note that the work opportunity cost is endogenous and is a function of the level of utility attained by the agent. In fact, under Assumption 1, it is non decreasing with this utility level. This implies that, under Assumption 1, among the set of Pareto optimal allocations, the aggregate employment level is highest at the *laissez-faire* equilibrium. A social objective that maximizes employment implicitly favors *laissez faire*.

Remark 2.1. The result sheds some light on an often asked policy issue. Consider a currently unemployed person, who collects benefits. Suppose that, if she does not take a job, the government or the unemployment agency will keep on paying the benefits. Should the agency financially help her finding a job, by using some of the money to subsidize the employer by reducing the wage cost? Should it make work pay, increasing earnings net of taxes and the financial incentives to work of the person? In a first best context, according to Theorem 1, the person should indeed be pushed to work by all means provided that her productivity is larger than her work opportunity cost, the difference being a gain for society. In the marginal case, this implies lifting all social contributions and taxes on work, while maintaining a subsidy equal to the unemployment benefits (no gain at all for the government!), so that after tax earnings just make up for the work opportunity cost.

2.3 Second best allocations and tax subsidy schemes

I now turn to more realistic second best situations, where the distribution of characteristics in the economy is common knowledge, but the individual agent's taste for work is private information and her productivity is only publicly known when she has a job. Second best allocations are typically implemented through a tax-benefit schedule which specifies the level of subsistence income when unemployed and relates the after tax income of the agent to her productivity when she works.

Formally I assume that agent a's productivity w is observed by the government *only when agent a works* (then it is her before-tax wage income), and that the unobservable individual characteristics y of the agent cannot be used to base the tax-subsidy scheme. The government, however, observes the characteristics x, and the taxes and subsidies can be made conditional on x. The government also knows the (typically non degenerate) distribution of unobservables y in the economy, conditional on (w, x). Under these assumptions, all allocations give an income guarantee $r(x)$ to the agents of characteristics x when unemployed (it cannot depend on w nor on y which are private information) and an income $R(w, x)$ to the workers (again independent of y).

These allocations can be obtained, without loss of generality, through a tax-subsidy schedule posted by the government. A tax-subsidy schedule is a couple $(r(x), D(w; x))$, where the subsistence revenue of the non worker is $r(x)$, $r(x) \geq 0$, and $R(w; x) = r(x) + D(w; x)$, $R(w; x) \geq 0$, is the income of the worker of productivity w. The *financial incentives to work* provided by the government policy are $D(w; x)$. I assume that, when they work, the agents reveal their true productivity w. If they have the possibility, without cost, to behave as agents of lower productivities, truthful revelation would only obtain under the condition that $D(w, x)$ be non decreasing in w, a condition worth keeping in mind when looking at the results.

Facing such a schedule, an agent a chooses either to work and receive $R(w; x)$ or not to work and receive $r(x)$. She decides to work when $\Delta(r(x); a) \leq D(w; x)$, with indifference in case of equality[5]. All the workers with productivity w and

[5]To simplify notations, I shall assume in the remainder of the paper that the agent chooses to work in the border case, when $\Delta(r(x); a) = D(w; x)$.

work opportunity cost strictly lower than $D(w;x)$ get a rent, compared to the unemployment situation, equal to $D(w,x) - \Delta(r(x);a)$. A natural way to look at an allocation in the economy is to stratify the agents according to their characteristics (w,x). In each stratum (w,x), there are two groups: the unemployed with $\Delta(r(x);a) > D(w,x)$, who receive an income $r(x)$, and the workers, with $\Delta(r(x);a) \leq D(w,x)$, who get an after-tax income $r(x)+D(w,x)$, and pay taxes equal to $w - D(w,x) - r(x)$.

Recall that $G_{r,w,x}$ is the c.d.f. of the distribution of work opportunity costs $\Delta(r;a)$ conditional on the agent productivity w

$$G_{r,w,x}(D) = \Pr\left(\Delta(r;a) \leq D \mid w,x\right).$$

The proportion of agents of type x with productivity w who work when the schedule is $(r(x), D(w;x))$ is $G_{r,w,x}(D(w;x))$. The pair $(r(x), D(w;x))$ is *feasible* when it satisfies

$$\int [w - D(w;x)]\mathbb{1}_{\Delta(r(x);a)\leq D(w;x)} \mathrm{d}F(a) = \int r(x)\mathrm{d}F(a),$$

which can be rewritten

$$\int [w - D(w;x)]G_{r,w,x}(D(w;x))\mathrm{d}\tilde{F}(w,x) = \int r(x)\mathrm{d}F(a), \tag{3}$$

where $\tilde{F}$ is the joint distribution of productivities and observable characteristics in the population. The left hand side of (3) is equal to the government revenue, which serves to finance a universal transfer r to everyone in the economy. The feasibility constraint is also equivalent to

$$\begin{aligned}\int [w - D(w;x) - r(x)]G_{r,w,x}(D(w;x))\mathrm{d}\tilde{F}(w,x) &= \\ &\int r(x)[1 - G_{r,w,x}(D(w;x))]\mathrm{d}F(a),\end{aligned}$$

where the left hand side denotes the taxes collected on workers which are equal to the transfers to the unemployed on the right hand side.

Any feasible tax-subsidy schedule yields an allocation $(c(a), s(a))$, which satisfies the standard incentive compatibility definition. In Choné and Laroque (2001), we show that the converse holds, and that there is no loss of generality in working with tax-subsidy schedules as above, compared with incentive compatible allocations.

Second best optima correspond to feasible incentive compatible allocations (or associated tax-subsidy schedules) which are not Pareto dominated by any other feasible incentive compatible allocations (or tax-subsidy schedule). The *laissez-faire* allocation, obtained with $r(x) = 0$ and $D(w, x) = w$ for all (w, x), is an optimum. More interestingly, the optima typically assign some positive subsistence income $r(x)$ to the unemployed in the economy, depending on the weight of the less well off in the social objective, financed through a tax $w - D(w, x) - r(x)$ on the employed.

2.4 The Laffer bound

At an optimum, taxes must not be too high (or equivalently, incentives too low), in order to keep the economy on the right side of the Laffer curve; otherwise decreasing the tax rates increases tax revenues, which can be used to make everyone better off. Indeed, given any a priori non-negative minimum income guarantee $r(x)$, there is a lower bound $d_r(w; x)$ on incentives $D(w; x)$, which can be described in terms of the fundamentals of the economy and whose properties will be studied in detail. I shall refer to this bound as the *Laffer bound*.

Theorem 2. *If the tax subsidy scheme* $(r(x), D(w; x))$ *is second best optimal, then*

$$D(w; x) \geq d_r(w; x) \text{ for all } w, x$$

where

$$d_r(w; x) = \sup \underset{D, D \leq w}{\operatorname{argmax}} (w - D) G_{r(x),w,x}(D). \tag{4}$$

Proof: the allocation satisfies, by construction:

$$c(a) = r(x) \mathbb{1}_{\Delta(r;a) > D(w(a);x)} + [r(x) + D(w(a); x)] \mathbb{1}_{\Delta(r;a) \leq D(w(a);x)}.$$

Using the utility index measuring equivalent consumption at work, it follows that:

$$\begin{aligned} \int u(c(a), s(a), a) dF(a) &= \int r(x) dF(a) \\ &\quad + \int \max[D(w(a); x), \Delta(r(x); a)] dF(a), \end{aligned}$$

or, equivalently, using the feasibility constraint (3):

$$\begin{aligned} \int u(c(a), s(a), a) dF(a) &= \int [w(a) - D(w(a); x)] G_{r,w(a),x}(D(w(a); x)) dF(a) \\ &\quad + \int \max[D(w(a); x), \Delta(r(x); a)] dF(a). \end{aligned}$$

The first term on the right hand side is government revenue from taxes used to fund the minimum income guarantee $r(x)$. The second term describes the surplus the agents get on top of $r(x)$, counted in utility units at work. It is equal to $\Delta(r(x); a)$ for the unemployed, and to $D(w(a); x)$ for the workers.

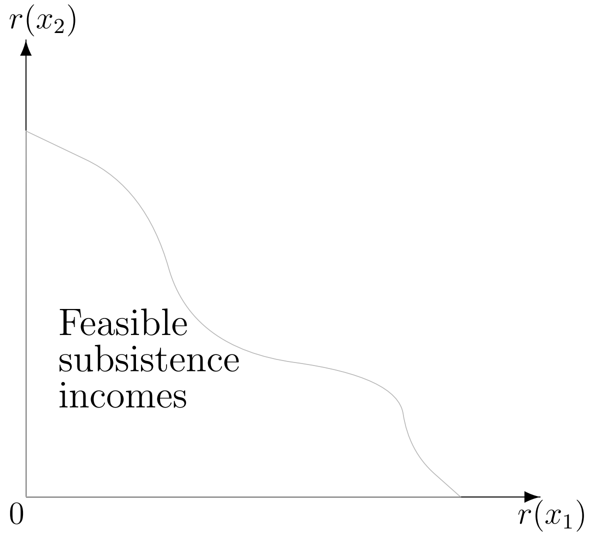

Figure 3: The set of feasible subsistence incomes when there are two observable types

Suppose that for some positive measure set of agents $D(w;x) < d_r(w;x)$. All these agents are better off if $D(w;x)$ is replaced with $d_r(w;x)$. Furthermore government revenue is larger, and the collected surplus can be redistributed to increase $r(x)$, thereby increasing everybody's utility, a contradiction. ■

2.5 Towards a characterization of second best optima

Theorem 2 paves the way towards a characterization of the second best tax schedules. It gives a necessary condition which has to be satisfied, given the level of subsistence income $r(x)$. I now first qualitatively describe the set of feasible subsistence incomes, before turning to second best tax schedules associated with these subsistence incomes.

2.5.1 Feasible subsistence incomes

Given the subsistence income or universal benefit $r(x)$, feasibility requires (3):

$$\int [w - D(w;x)] G_{r,w,x}(D(w;x)) \mathrm{d}\tilde{F}(w,x) = \int r(x) \mathrm{d}F(a).$$

Government revenue raised on the workers must cover the cost of the universal benefit. Denote by $K_r(w,x)$ the (maximal) revenue collected at the Laffer bound:

$$K_r(w,x) = \max_{D, D \leq w} (w - D) G_{r(x),w,x}(D) = (w - d_r(w,x)) G_{r(x),w,x}(d_r(w,x)).$$

It is easy to show[6] that, under Assumption 1, $K_r(w,x)$ is a continuous non increasing function of $r(x)$. By definition, if a universal benefit $r(x)$ is feasible, then

$$\int K_r(w,x)\mathrm{d}\tilde{F}(w,x) \geq \int r(x)\mathrm{d}F(a),$$

and conversely, if the above inequality holds, one can associate a feasible tax schedule to $r(x)$ satisfying the government budget constraint at equality, by sufficiently increasing $D(w,x)$ above $d_r(w,x)$.

To get a more precise description of the feasible r's, consider first the situation where the set of characteristics x is a singleton, so that the subsistence income is uniform in the population. Since K_r is continuous and non decreasing, the equation in r

$$r - \int K_r(w)\mathrm{d}\tilde{F}(w) = 0$$

has a unique positive solution $\bar{r}$ and all r in $[0,\bar{r}]$ are feasible. When r is equal to the upper bound of the feasible range $\bar{r}$, there is a unique associated feasible tax schedule, which is equal to the Laffer bound $D(w) = d_r(w)$. For smaller values of r, there are a lot of feasible taxes, which have to satisfy

$$\int [w - D(w)]G_{r,w}(D(w))\mathrm{d}\tilde{F}(w) = r,$$

even if one restricts the attention to $D(w)$'s which are larger than the Laffer bound.

The case where there are a number of observable characteristics is similar. Consider for instance $k = 1,\ldots,K$ values of x. Pick any (not too large) $r(x_2),\ldots,r(x_K)$ and find r_1 as the (at most unique) solution of the equation

$$r_1 - \int K_{r_1}(w,x_1)\mathrm{d}\tilde{F}(w,x_1) + \sum_{k=2}^{K}\left[r(x_k) - \int K_{r_k}(w,x_k)\mathrm{d}\tilde{F}(w,x_k)\right] = 0.$$

When $(r(x_2),\ldots,r(x_K))$ varies, this describes a surface in the positive orthant of dimension K, and under minimal regularity conditions the feasible set is made of all the points below this surface (see Figure 3 for $K = 2$). For points on the boundary, the only feasible tax schedule is the Laffer bound. For points inside, there are many possible taxes, and I now single out those that are second best efficient.

2.5.2 Second best tax schedules

The reason why a schedule with incentives above the Laffer bound might not be optimal is similar to what makes schedules below the bound non optimal: it may

[6] The proof needs some care since G may not be continuous, when the distribution has mass points. See Lemma C.2 in the appendix to Choné and Laroque (2001) for a formal argument.

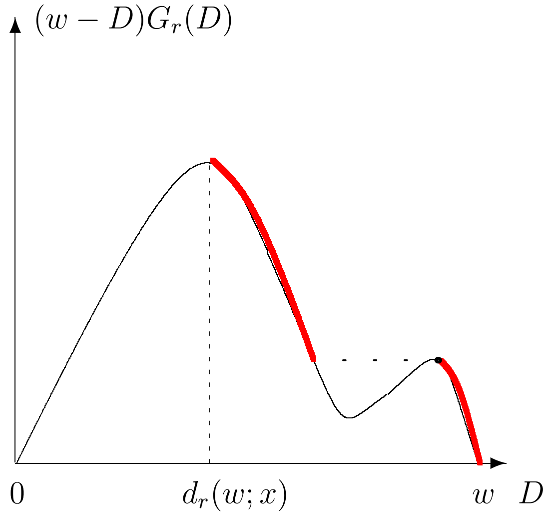

Figure 4: The Laffer bound and overtaxation

be the case that larger incentives, i.e. lower taxes, yield the same government revenue while making all the workers better off (see Figure 4). This motivates the following definition:

Definition 1. *The set of agents of characteristics (w, x) is* overtaxed *at the tax schedule $(r(x), D(w; x))$ if there is some d, $d > D(w; x)$, such that*

$$(w - d)G_{r(x),w,x}(d) \geq (w - D(w; x))G_{r(x),w,x}(D(w; x)).$$

The following result shows that if one limits the attention to feasible tax schemes such that the agents' incentives to work $D(w, x)$ are smaller than their productivities w, then they are second best efficient if and only if they do not overtax. Situations where $D(w, x)$ is larger than w are discussed below in Remark 2.3.

Theorem 3. *Consider a feasible tax schedule $(r(x), D(w; x))$ such that $D(w, x) \leq w$ for all (w, x). Under Assumption 1, the tax schedule $(r(x), D(w; x))$ supports a second best optimal allocation if and only if (almost) no category is overtaxed at $(r(x), D(w; x))$.*

It follows from Theorem 3 that the Laffer bound is second best efficient. When subsistence incomes are on the feasibility frontier, Laffer taxes are the only feasible taxes. The further subsistence incomes away from the boundary, or the closer to zero, the larger the distance of the tax schemes away from the Laffer bound and the less constrained they are. The next section studies in detail the properties of the Laffer bound. But away from this bound, efficiency imposes essentially no constraint on the local shape of the tax scheme: it allows, for instance, after tax incomes whose derivative with respect to w is larger than one (negative apparent marginal tax rates).

Proof: the proof of necessity follows the same lines as that of Theorem 2 and is omitted for the sake of brevity. I give the sufficiency argument for an economy with a single type x, leaving the generalization to the interested reader. The feasibility constraint (3) is:

$$r = \int [w - D(w)] G_{r,w}(D(w)) \mathrm{d}\tilde{F}(w).$$

Consider a tax scheme $(r, D(w))$ satisfying the assumptions of the Theorem and let $(r', D'(w))$ be a feasible alternative. Suppose, by contradiction, that the allocation associated with $(r', D'(w))$ strictly Pareto dominates that which follows from $(r, D(w))$.

If $r' = r$, Pareto domination implies that $D'(w) \geq D(w)$ for all w, with a strict inequality for a non negligible set of wages. This is not compatible with the feasibility constraint, under the assumption of no overtax at r.

Suppose now that $r' > r$. Define a partition of the set of productivities as follows:

$$\begin{aligned} W_1 &= \{w | D'(w) \geq w\}, \\ W_2 &= \{w | w > D'(w) \geq D(w)\}, \\ W_3 &= \{w | D(w) > D'(w)\}. \end{aligned}$$

The feasibility constraint (3) can be written:

$$\begin{aligned} r = & \int_{W_1} [w - D(w)] G_{r,w}(D(w)) \mathrm{d}\tilde{F}(w) + \int_{W_2} [w - D(w)] G_{r,w}(D(w)) \mathrm{d}\tilde{F}(w) + \\ & \int_{W_3} [w - D(w)] G_{r,w}(D(w)) \mathrm{d}\tilde{F}(w). \end{aligned}$$

The proof compares each of the three terms on the right hand side with the corresponding terms for the prime allocation, leading to a contradiction. First by assumption on $D(w)$ and definition of W_1:

$$\int_{W_1} [w - D(w)] G_{r,w}(D(w)) \mathrm{d}\tilde{F}(w) \geq 0 \geq \int_{W_1} [w - D'(w)] G_{r',w}(D'(w)) \mathrm{d}\tilde{F}(w).$$

Second, by the no overtax property, given the minimum income guarantee r:

$$\int_{W_2} [w - D(w)] G_{r,w}(D(w)) \mathrm{d}\tilde{F}(w) \geq \int_{W_2} [w - D'(w)] G_{r,w}(D'(w)) \mathrm{d}\tilde{F}(w),$$

so that, using assumption 1, $G_{r',w}(D) \leq G_{r,w}(D)$, and the fact that $w \geq D'(w)$ on W_2

$$\int_{W_2} [w - D(w)] G_{r,w}(D(w)) \mathrm{d}\tilde{F}(w) \geq \int_{W_2} [w - D'(w)] G_{r',w}(D'(w)) \mathrm{d}\tilde{F}(w).$$

Finally, for the third term, start with assumption 1 to get

$$\int_{W_3} [w - D(w)] G_{r,w}(D(w)) \mathrm{d}\tilde{F}(w) \geq \int_{W_3} [w - D(w)] G_{r',w}(D(w)) \mathrm{d}\tilde{F}(w).$$

From Pareto domination, the workers in W_3 must be as well off in the prime allocation as in the reference allocation, i.e. $r' + D'(w) \geq r + D(w)$. Substituting for $D(w)$ in the right hand side, one gets:

$$\begin{aligned}\int_{W_3} [w - D(w)] G_{r,w}(D(w)) \mathrm{d}\tilde{F}(w) \quad &\geq \quad \int_{W_3} [w - D'(w)] G_{r',w}(D(w)) \mathrm{d}\tilde{F}(w) + \\ &\quad (r - r') \int_{W_3} G_{r',w}(D(w)) \mathrm{d}\tilde{F}(w),\end{aligned}$$

and, using the fact that $D(w) > D'(w)$ on W_3:

$$\begin{aligned}\int_{W_3} [w - D(w)] G_{r,w}(D(w)) \mathrm{d}\tilde{F}(w) \quad &\geq \quad \int_{W_3} [w - D'(w)] G_{r',w}(D'(w)) \mathrm{d}\tilde{F}(w) + \\ &\quad (r - r') \int_{W_3} G_{r',w}(D(w)) \mathrm{d}\tilde{F}(w).\end{aligned}$$

Coming back to the initial equality, using feasibility at the prime allocation, (3) becomes:

$$\left(1 - \int_{W_3} G_{r',w}(D(w)) \mathrm{d}\tilde{F}(w)\right)(r - r') \geq 0,$$

which implies

$$1 = \int_{W_3} G_{r',w}(D(w)) \mathrm{d}\tilde{F}(w) = \int_{W_3} G_{r,w}(D(w)) \mathrm{d}\tilde{F}(w).$$

The set W_3 is of full measure, so that $D(w) > D'(w)$ for all w, and everyone works at the reference allocation. Now let A' be the set of agents that work at the prime allocation and $\bar{A}'$ its complement. Taking the difference of the two feasibility constraints yields:

$$\int_{A'} [r + D(w) - r' - D'(w)] dF(a) - \int_{\bar{A}'} [r' + w - D(w) - r] dF(a) = 0.$$

But the first term on the left hand side is non positive by Pareto domination (strictly negative if A' is the full set) as well as the second term ($r' > r$ and $w \geq D(w)$), which yields the desired contradiction.

Suppose finally that $r' < r$. To be better off in the prime allocation than in the original one, every one has to get a utility level at least equal to $r + \Delta(r; a)$ and therefore has to work. Moreover, $r' + D'(w)$ is at least as large as $r + D(w)$. It follows that $\bar{D}(w) = D'(w) + r' - r$ is as large as $D(w)$. By construction, when confronted with incentives $\bar{D}(w)$ at subsistence income r, everyone wants to work. The government revenue in this hypothetical situation is:

$$\int [w - \bar{D}(w)] \mathrm{d}\tilde{F}(w) = \int [w - D'(w) - r' + r] \mathrm{d}\tilde{F}(w).$$

The feasibility of the prime allocation implies

$$\int [w - D'(w)] \mathrm{d}\tilde{F}(w) = r',$$

so that

$$\int [w - \bar{D}(w)] \mathrm{d}\tilde{F}(w) = r.$$

The allocation $(r, \bar{D}(w))$ would be feasible! But this contradicts the assumption on government revenue, since $\bar{D}(w) \geq D(w)$, with some strictly positive inequalities. ■

There is a natural assumption under which there is no overtax at $(r(x), D(w; x))$ whenever $D(w; x)$ is larger than the Laffer bound at $r(x)$. The assumption simply states that government revenue decreases with the level of incentives above the Laffer bound:

Assumption 2. *For all* (r, x, w) *government revenue*

$$(w - D) G_{r,w,x}(D)$$

is decreasing in D *for* D *larger than* $d_r(w; x)$.

Assumption 2 has an intuitive content, but is restrictive. It is satisfied when the distribution of work opportunity costs $G(D)$ is log-concave (see Theorem 5 for further properties of the Laffer tax schedule when $G(D)$ is log-concave), a property which may be easier to check than the no overtax property in applications.

Remark 2.2. In the Ramsey tradition, see also Saez (2001) for a recent application, one can rephrase the previous results in terms of (average) tax rates[7]. Define the tax rate τ faced by an agent a of productivity w, with financial incentives to work D, as

$$\tau = 1 - \frac{D}{w}.$$

[7] I owe this remark to an anonymous referee.

Note that this definition is in terms of incentives: the common tax rate links before tax to after tax income and is equal to $1-(r+D)/w$, but is not relevant here. When the function G is differentiable, the first order condition satisfied by the Laffer tax is

$$-G_{r,w,x}(d)+(w-d)G'_{r,w,x}(d)=0,$$

or, equivalently when d is different from zero,

$$\frac{w}{d}-1=\frac{G_{r,w,x}(d)}{dG'_{r,w,x}(d)},$$

that is

$$\frac{\tau_L}{1-\tau_L}=\frac{1}{\varepsilon_{r,w,x}(d)},$$

where τ_L is the Laffer tax rate and $\varepsilon_{r,w,x}(d)$ is the elasticity of labor supply with respect to the financial incentives to work. Under Assumption 2, Theorem 3 then says, that *if one only considers allocations such that everybody's tax rates are positive ($D \le w$)*, they are second best Pareto optimal if and only if taxes are smaller than their Laffer bound, $\tau \le \tau_L$. Allocations with $D > w$ or $\tau < 0$ for some w, on the right side of the Laffer curve (satisfying Theorem 2), are not always second best optimal. Such allocations are not to be ruled out, especially when the economy has agents with very low productivities (see next remark).

Remark 2.3. In a first best optimum, nobody works when her productivity is smaller than her work opportunity cost. Theorem 3 only deals with situations where the incentives to work $D(w;x)$ are smaller than w, so that nobody with a work opportunity cost larger than her productivity will choose to work. However, contrary to the first best, there are second best allocations, not covered by Theorem 3, where the planner puts a lot of weight on the welfare of agents of characteristics $(w;x)$ and finds it worthwhile to set a value of $D(w;x)$ larger than w.

Here is a simple numerical example, with (too) many 'working poors'. The agents' productivities take two values, either $w_0=0$ or $w_1=3$, with equal probabilities. The agents with zero productivities have a work opportunity cost Δ_0 uniformly distributed on the segment $[0,2]$, while the high productivities all have $\Delta_1=0$. Any feasible allocation $(r,D(w))$ satisfies

$$2r+\min\left(\frac{D(w_0)}{2},1\right)D(w_0)+D(w_1)=w_1,$$

and the Laffer bound, here independent of r, is $d(w_0)=d(w_1)=0$. Then the allocation $r=0$, $D(w_0)=2$ and $D(w_1)=1$ is feasible and taxes less than the

Laffer tax, but it is Pareto dominated by $r' = 1$, $D'(w_0) = 1$ and $D'(w_1) = 1/2$, which itself is a second best optimum[8].

A close look at the proof of Theorem 3 shows that an allocation with $D(w) > w$ for some w which is on the right side of the Laffer curve can only be Pareto dominated by another allocation with a larger subsistence income. In practice, actual tax benefit schedules do not seem to have incentives to work larger than productivity. For instance, on Figure 1, disposable income is always below the line of slope one that originates at its intercept.

Example: As an illustration, consider a simple numerical economy, where there is a single type of agent (X is degenerate) and the distribution of work opportunity costs does not depend on r, nor on productivity w. Suppose that the work opportunity costs and productivities both take their values in $[1, +\infty)$ with c.d.f. respectively

$$\tilde{F}(w) = 1 - \frac{1}{w^2} \qquad G(D) = 1 - \frac{1}{D}.$$

The Laffer bound maximizes $(w - D)G(D)$, and is $d(w) = \sqrt{w}$. Therefore:

$$\bar{r} = \int_1^{+\infty} (w - \sqrt{w})G(\sqrt{w})\mathrm{d}\tilde{F}(w) = 1/3.$$

From Theorem 3, for r smaller than $\bar{r}$, any incentive scheme $D(w)$, $w \geq D(w) \geq \sqrt{w}$, which is feasible,

$$r = \int_1^{+\infty} (w - D(w))\mathrm{d}\tilde{F}(w),$$

yields a second best optimum. For instance keep $D(w)$ equal to $\sqrt{w}$ except on some interval $[w_0, w_1]$ where $D(w) = \sqrt{w_0} + 2(w - w_0)$ with w_1 such that

$$\bar{r} - r = \int_{w_0}^{w_1} (D(w) - \sqrt{w})\mathrm{d}\tilde{F}(w),$$

[8]To show that the prime allocation dominates the initial one, note first that all the high productivity agents work and are better off, with 3/2 units of good instead of 1. The low productivity agents of work opportunity cost larger than 1 do not work and receive 1 unit of good, while they previously worked with a consumption net of work cost equal to $2 - \Delta_0 \leq 1$. The other low productivity agents still work and get 2 units of good as before. Furthermore, the prime allocation is second best optimum. Indeed any allocation that Pareto dominates it must give at least 3/2 units of good to the high productivity types, so that by feasibility $r + D(w_0)^2/2 \leq 3/2$. The low productivity types of zero work opportunity cost get $r + D(w_0)$, which must be larger than 2. But the maximum of $r + D$ over the constraint $r + D^2/2 \leq 3/2$ is obtained at $D = 1$. Finally, it also is very easy, and may be relevant, to present examples where the work opportunity cost varies with r. For instance, modify the distribution of opportunity costs of the low productivity workers in the above economy. Suppose that it is concentrated on $\Delta_0 = 1$ when r is equal to zero, but that when r equals 1, perhaps because of increasing returns in home production, it is concentrated on 3. Then the allocation $r'' = 1$, $D''(w_0) = 2$ and $D''(w_1) = 1$ Pareto dominates the reference allocation: the agents of low productivity enjoy not to work, when they have the subsistence income of 1.

which shows the possibility of negative income taxes on $[w_0, w_1]$. The example can of course be adapted to preserve monotonicity of after tax income.

Remark 2.4. The above result goes well with the examples provided by Diamond (1980), who studies an extensive model as here, under a utilitarian criterion. Indeed, it seems that very few restrictions can be expected to hold on the shape of the optimal tax schemes, except perhaps for very special forms of the utilitarian criteria. By contrast, the intensive model (Stiglitz (1982)) leads to rather sharp conclusions: for instance, the marginal tax rate has to stay between zero and one. It would be of interest to characterize the set of second best allocations in an economy with both intensive and extensive features.

2.6 Qualitative analysis of the Laffer tax scheme[9]

The previous results emphasize the importance of the Laffer tax, which maximizes government revenue. It turns out that this maximization has an interesting structure which is most transparent when the distribution of work opportunity costs is independent of the productivity of the agents:

Assumption 3. *The conditional distribution of work opportunity costs $G_{r,w,x}(.)$ is independent of w.*

The basic structure of the problem

Under Assumption 3 the optimization problem has two important features: the objective is linear with respect to productivity w and it depends in a simple way on the distribution of work opportunity costs.

Theorem 4. *Under Assumptions 1 and 3, we have*

1. *the maximal revenue $K_r(w,x) = (w - d_r(w;x))G_{r,x}(d_r(w;x))$ raised by the government is a non decreasing convex positive function of w, of slope at most equal to 1.*

2. *$d_r(w;x)$ is a nondecreasing function of w. The proportion of agents of type x and productivity w at work, $G_{r,x}(d_r(w;x))$, is also nondecreasing in w.*

Proof: From Theorem 2, $K_r(w,x)$ is the supremum of the set of linear mappings $(w-d)G_r(d)$, where d is any real number. It is positive ($d = w$ is possible), convex as the supremum of convex functions. $G_{r,x}(d_r(w;x))$ is a gradient of $K(w)$, whose slope cannot thus exceed 1. Convexity implies that the gradient is nondecreasing, which implies that $G_{r,x}(d_r(w;x))$ is nondecreasing in w, and $d_r(w;x)$ as well. ∎

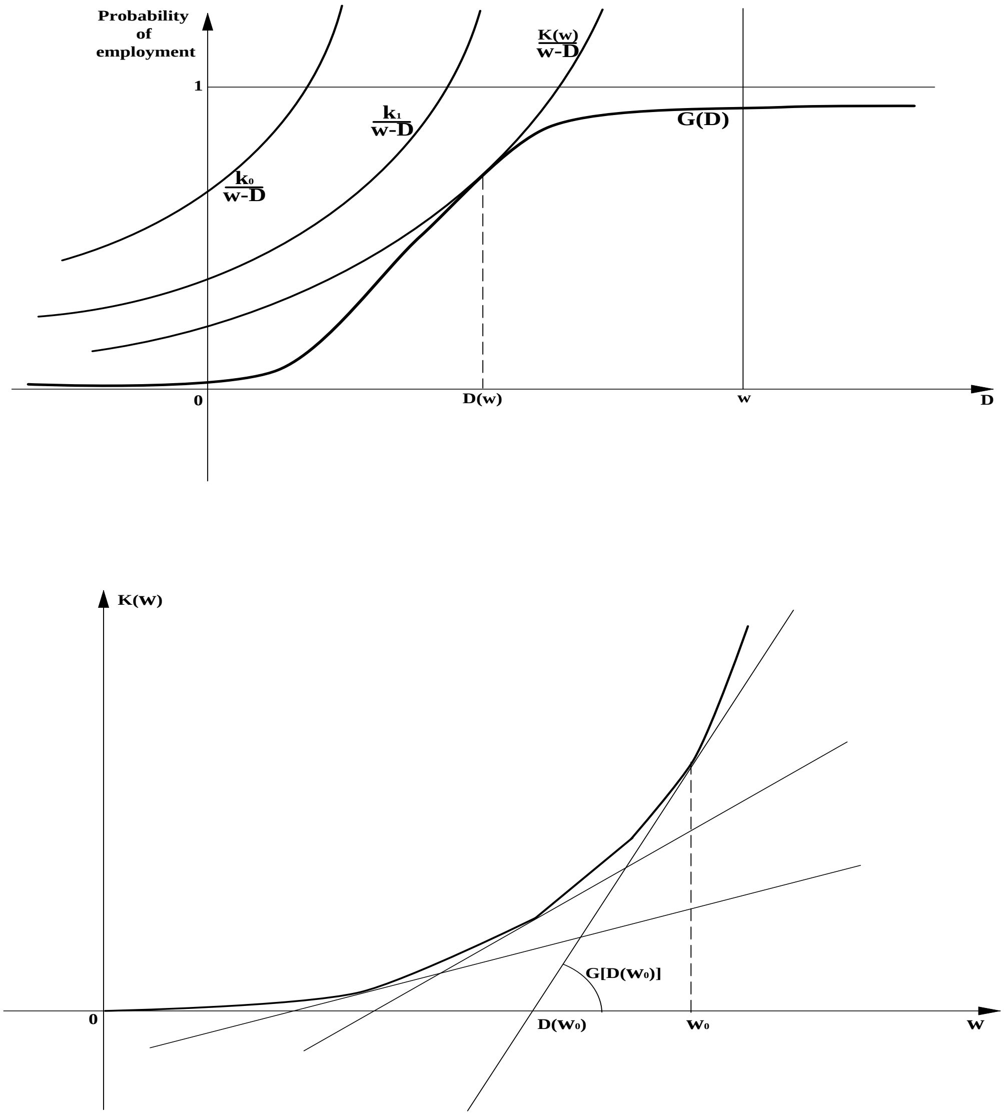

Figure 5: The optimization program

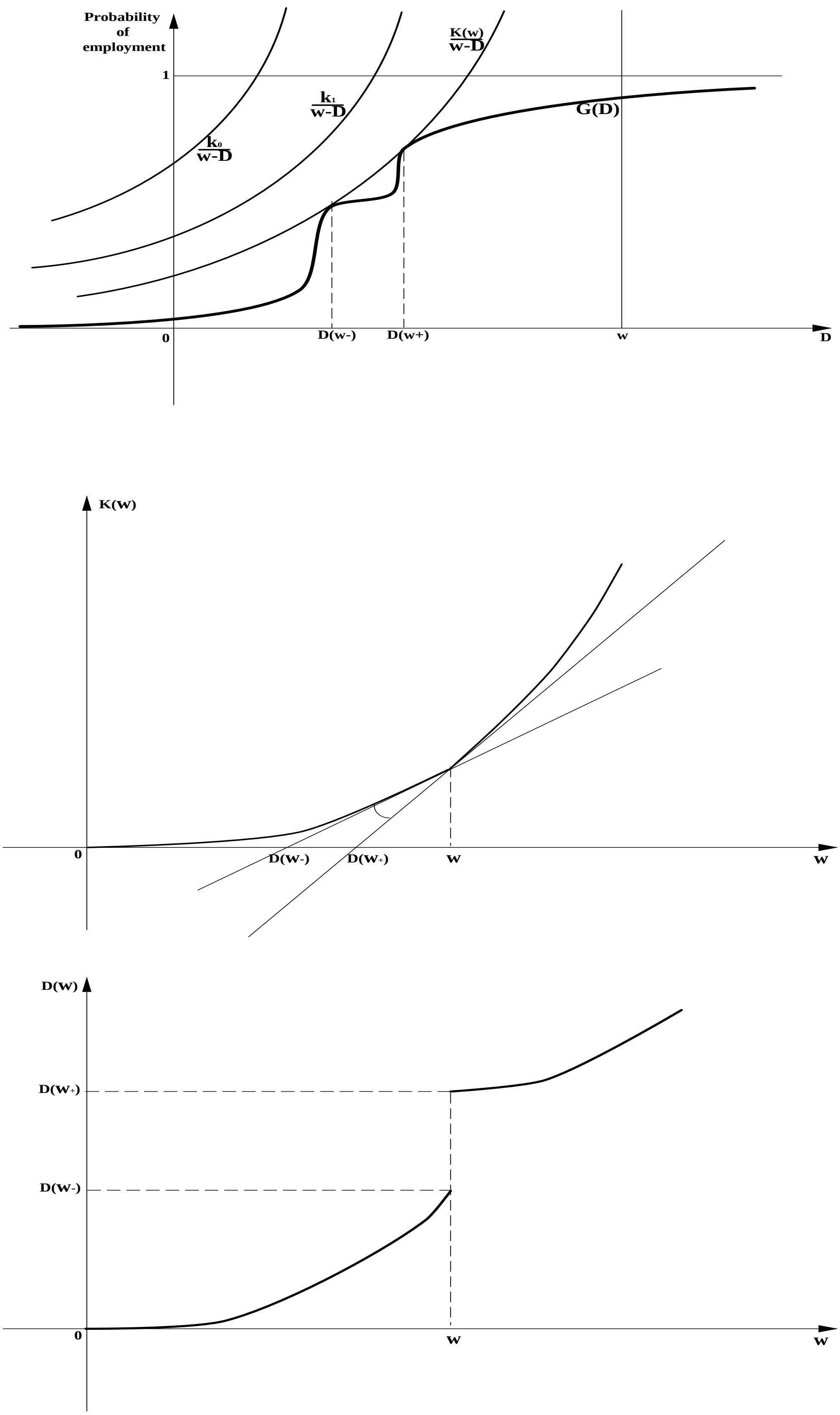

Figure 6: Discontinuity of the tax scheme

The theorem shows that, under Assumption 3, the marginal tax rates $1 - d'_r(.;x)$ are less than or equal to 1. The fact that $d_r(w;x)$ is nondecreasing in w implies that it would not be in the interest of an agent to announce a productivity lower than the truth, if this were allowed. The Laffer tax schedule is incentive proof to the mimicking of agents with lower productivities.

A graphical representation, where for simplicity the characteristics x are omitted, helps to understand the structure of the problem. On the top panel of Figure 5, the c.d.f. $G_r(D)$ is plotted: if D is selected by the government, $G_r(D)$ is the proportion of agents that are willing to work. For a given value of w, the problem is to find the maximum value of k such that $k/(w - D)$ intersects the graph of the c.d.f.. Therefore, for a given w, I draw a bunch of isoquants of the form $k/(w - D)$, all arcs of hyperbolas whose asymptotes are the negative D axis and the vertical line of abscissa w. The solution is at the highest isoquant which is tangent to the c.d.f.. When w increases, the hyperbolas translate to the right, so that both $d_r(w)$ and $K_r(w)$ increase. It is also of interest to locate the *laissez-faire* allocation on the graph. For r equal to zero, all the agents with a work opportunity cost larger than w, the top of the distribution, do not work and get 0. The rest of the population works and enjoys a surplus from work equal to the horizontal distance between the graph of G and the vertical line of abscissa w.

The point-wise optimization program, for a specific value of w, need not be well behaved. However, the overall optimization is simple, as shown on the bottom panel of Figure 5 drawn in the plane $(w, K(w))$. The maximization involves taking the upper envelope $K(w)$ of a set of straight lines of equation $(w - D)G_r(D)$, when D varies. The typical line intersects the w axis at D, and has slope $G_r(D)$, a number between 0 and 1. The function $K_r(w)$ is increasing convex (and therefore continuous), and has a slope everywhere smaller than 1.

When do Laffer taxes have negative marginal tax rates?

The top panel of Figure 6 shows a situation where there are two tangency points. For this particular value of w, both $D(w-)$ and $D(w+)$ maximize government revenue. This results in an upward discontinuity of $d_r(.)$, which tends towards $D(w-)$ when its argument approaches w from the left, while $d_r(w) = D(w+)$. The Laffer tax scheme exhibits an infinite negative marginal tax rate[10] at w.

Two remarks are worth making at this stage. First, such discontinuities have nothing pathological: they will occur as soon as the c.d.f. has pieces that are

[9] This section is based on the work of Choné and Laroque (2001).

[10] Here, the 'marginal tax rate' denotes the derivative of taxes with respect to w. In the extensive model, labor supply decisions depend on the average tax rate, while in the intensive model they are directly linked with the marginal tax rate (see Remark 2.2).

flatter than the arc of hyperbola going through them, for instance for discrete distributions. Second, I have represented the extreme case of an infinite negative tax rate. This should not induce the reader to believe that *finite* negative marginal tax rates are impossible. Actually, essentially every nondecreasing schedule is the Laffer schedule associated with a well chosen distribution of work opportunity costs (see the study of the inverse problem in Choné and Laroque (2001) which makes this assertion precise).

To understand intuitively why negative marginal tax rates can help maximize government revenue, suppose there is an accumulation of agents with work opportunity cost close to d (d being known to the planner). Recall that work opportunity cost is unobserved in the second best environment: the only available screening variable is w. For small w, it is too costly to put these agents to work while it is optimal to do so for large w. If the distribution of opportunity costs is very concentrated around d, the Laffer schedule is such that the incentives strongly increase ($D'(w) > 1$) precisely at the point w such that $D(w) = d$.

There exists a simple regularity assumption on the distribution of work opportunity costs that guarantees that the Laffer tax schedule always has positive marginal tax rates. In particular, this assumption rules out mass points in the distribution G.

Proposition 5. *When G is log concave, the Laffer marginal tax rate is everywhere nonnegative.*

Proof The problem (4) can be rewritten $\max_{D \leq w} \ln(w-D)+H(D)$, with $H(D) = \ln G$. Since H is concave, the function $D \to \ln(w - D) + H(D)$ is strictly concave and has a unique maximum, characterized by the first order conditions[11]

$$H'(D) = \frac{1}{w - D}.$$

Since D is nondecreasing and H' is nonincreasing, it follows that $w - D(w)$ increases in w, which gives the result. ■

3 An empirical illustration

Theory essentially says that any schedule with taxes less than the Laffer bound is second best optimal. The aim of this section is to illustrate the theory through a practical exercise. To compute the Laffer bound, I first posit a labor supply model. I discuss the assumptions that allow to recover from the data the joint distribution of productivities and work opportunity costs. The model then is

[11] When G has a kink, the first order condition is that 0 is in the gradient of $\ln(w-D)+H(D)$.

estimated, with a flexible non parametric distribution for the unobserved part of the work opportunity costs. After estimation, one obtains a joint distribution of (w, Δ), which depends on all the (exogenous) variables that are observable by the econometrician. The final step is to evaluate the Laffer tax scheme, based on the (assumed) subset of variables that are observable and verifiable by the tax authorities.

3.1 A labor supply model

In the remainder of the paper, I shall rely on a model developed on French data in Laroque and Salanié (2002). The model abstracts from a number of important features of real life, and the results below should be considered illustrative. The model is static and is applied to women who either do not work or have a full time job, and are between 25 and 49 years old[12]. It takes into account the high level of the minimum wage in France. The structure of the model is as follows. The typical woman's productivity satisfies :

$$\ln w = X\alpha + \sigma_\varepsilon \varepsilon, \tag{5}$$

where X includes age at end of studies and its square, work experience and its square and diploma in six categories. A woman has a job if the three following inequalities are satisfied:

1. Her productivity is higher than the cost to an employer of the minimum wage

$$w \geq w_{\min},$$

2. She is not subject to frictional or keynesian unemployment

$$\nu \leq P_k(Y\beta), \tag{6}$$

   where ν is uniformly distributed in the interval $[0, 1]$, and $-\ln(P_k)$ is of the form $\beta_0^2 + \beta_1^2(\text{age} - 25)$;

3. Finally, she is willing to work, i.e.

$$R(w) \geq R(0) + Z\gamma + \rho\varepsilon + \sigma_\eta \eta. \tag{7}$$

   Here $R(.)$ is a known highly non linear function: $R(w)$ is the net after tax and subsidies income of the household when the cost of labor to the employer is equal to w. The variables Z include the out of work income $R(0)$ itself, denoted r in the preceding section, as well as the family composition (presence of a spouse, number of children by age range).

[12]The same model has been estimated for various subsets of the French population of working age (see e.g. Laroque and Salanié (2000)). For lack of information on their incomes, households with a self employed person are excluded from the analysis. The civil servants who have tenure also are excluded from the sample under study.

The unobserved heterogeneity is described by the three variables (ε, ν, η), which are assumed to be independently distributed.

From (7), the work opportunity cost of the theory, Δ, is equal to $Z\gamma + \rho\varepsilon + \sigma_\eta \eta$, while productivity is given by (5). The model, if it can be reliably estimated on a representative sample of the population, gives a joint distribution of (w, Δ), allowing to compute the Laffer taxes.

3.2 Semiparametric identification and estimation

The structural model depends heavily on the choice of exogenous variables, on the assumed functional (mostly linear) forms and on the design of the unobserved heterogeneity. I take these modelling choices as given[13]. Under these restrictions, I informally argue below that the parameters of the model are identified as well as the distributions of the unobserved heterogeneity (semiparametric identification). Indeed it is important for my purpose that the shape of the distribution of η comes from the data, and not be chosen by the econometrician: in the simulations below, the standard deviation of the estimated work opportunity cost Δ is of the order of 1150 euros, while the standard error of the unobserved heterogeneity term σ_η amounts to 850 euros. Unfortunately the model is complicated: I follow a piecemeal approach, using intuitions from results in the literature obtained in simpler setups for models with either one or two equations.

3.2.1 Identification

1. *Wage equation and the minimum wage.* The cost of the minimum wage relative to the overall wage distribution appears to be high in France, by comparison with other developed countries. This creates difficulties to identify the distribution of ε in the wage equation, in the low wage region (see Meyer and Wise (1983a) or Meyer and Wise (1983b)). Semiparametric identification here relies on the assumption that the distribution does not depend on the diploma, a variable in X, and that the more skilled agents have a wage distribution with support above the minimum wage.

2. *Frictional or keynesian unemployment.* The specification of this component of unemployment is ad hoc in the model. Exclusion restrictions make the distribution identifiable. Indeed, the only variable in Y is age. The exogenous variables which determine productivity (age at end of school, diplomas) are assumed not to be part of Y. Then, provided long studies high diplomas and no children characterize a set of persons who want

[13]There is a large literature that discusses the possibility of recovering structural parameters in (dynamic) labor supply models, see e.g. Blundell, Duncan, and Meghir (1998). This is outside the scope of this limited illustration and should be the subject of further research.

Probability	Location	Scale
0.81	0.00	1.06
0.15	0.15	0.00
0.04	0.00	3.41

Location: thousands of euros

Table 1: Mixture of logistic variables

to work and are not barred by the level of the minimum wage, their only reasons for being unemployed are frictional or keynesian, which allows to identify the distribution of ν.

3. *Participation equation.* The coefficients α and γ of the labor demand and labor supply equations are identified from exclusion restrictions: the diplomas appear in labor demand, not in labor supply, while family composition, spouse income and the tax scheme are determinants of labor supply, not of labor demand. The issue of interest is whether the distribution of η can be recovered from the data. By analogy with the familiar analysis of single index models (see e.g. Horowitz (1998)), everything else being given, the distribution of η in equation (7) is identified under location and scale normalizations, since there is a continuous variable on the right hand side of (7) which varies independently of the left hand side. Holding w fixed, if there are many observations with varying income of the spouse leading to continuous variation of $Z\gamma$, the corresponding proportions of participants in the labour market will give an estimate of the c.d.f. of $\sigma_\eta \eta$.

The above arguments discuss in turn the identification of each of the distributions of ε, η and ν, the other two being given. I do not know whether the joint distribution of (ε, η, ν) is identified. In any case, I limit the attention here to the study of the distribution of η, maintaining the assumption that ν and ε are independently distributed as uniform on $[0, 1]$ and standard normal[14].

3.2.2 Estimation of the distribution of η

After having tried several estimation techniques, which included non parametric adaptive estimation and all yielded similar results, the simplest and most reliable procedure turns out to be a maximum likelihood procedure, where a flexible functional form is used for the c.d.f. of η. The model is estimated on a cross section of 13 837 women in 1999. Starting from an initial specification where the

[14]It may be of some comfort to know that, as far as ε is concerned, normality is not rejected by the data when ν is uniform on $[0, 1]$ and η is logistic (see Laroque and Salanié (2002) for a test against a mixture of normals).

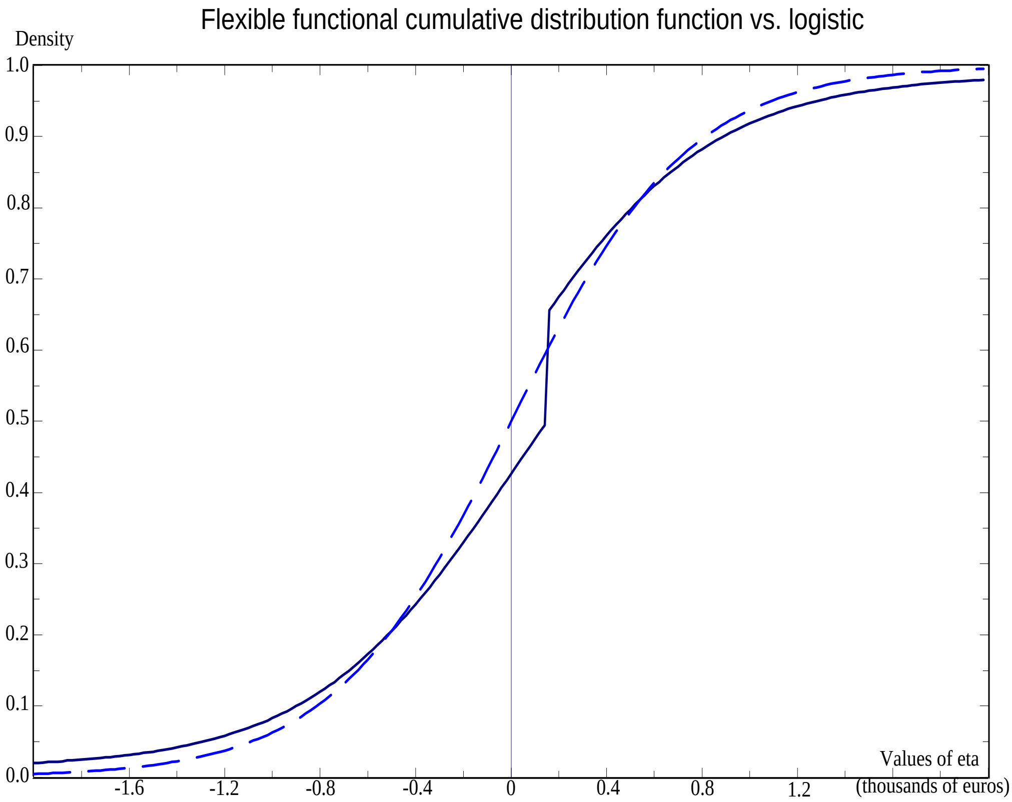

Figure 7: Flexible functional form

distribution of η is logistic, I fix the location parameter (constant term) and the scale parameter σ_η and allow the random variable η to be a mixture of (three) logistics, instead of a single standard one. More precisely, I replace the c.d.f.

$$\frac{1}{1+\exp(-x)}$$

with

$$\frac{p_1}{1+\exp(-\frac{x-\mu_1}{\sigma_1})} + \frac{p_2}{1+\exp(-\frac{x-\mu_2}{\sigma_2})} + \frac{p_3}{1+\exp(-\frac{x-\mu_3}{\sigma_3})}.$$

In other words, instead of assuming η to be a standard logistic with mean 0 and standard error $\pi/\sqrt{3}$, I assume that with probability p_i it is a logistic of mean μ_i (units: thousand euros) and standard error $\sigma_i\pi/\sqrt{3}$, for i equal to 1, 2 and 3, with $p_3 = 1 - p_1 - p_2$. This introduces eight extra parameters (three μ's and σ's and two probabilities), whose values are reported in Table 1. Most of the weight (more than 80%) is put on a distribution close to the initial standard logistic. But 15% goes to a Dirac mass, at a small positive abscissa (150 euros) and 4% corresponds to a fat tail, with a standard error 3.4 larger than the standard logistic. The mixture parameters globally are statistically significant with a p-value of 1.2%, since the log-likelihood function is increased from -7902.1 to -7892.3, by 9.8 points, for eight degrees of freedom. The associated c.d.f. is shown on Figure 7.

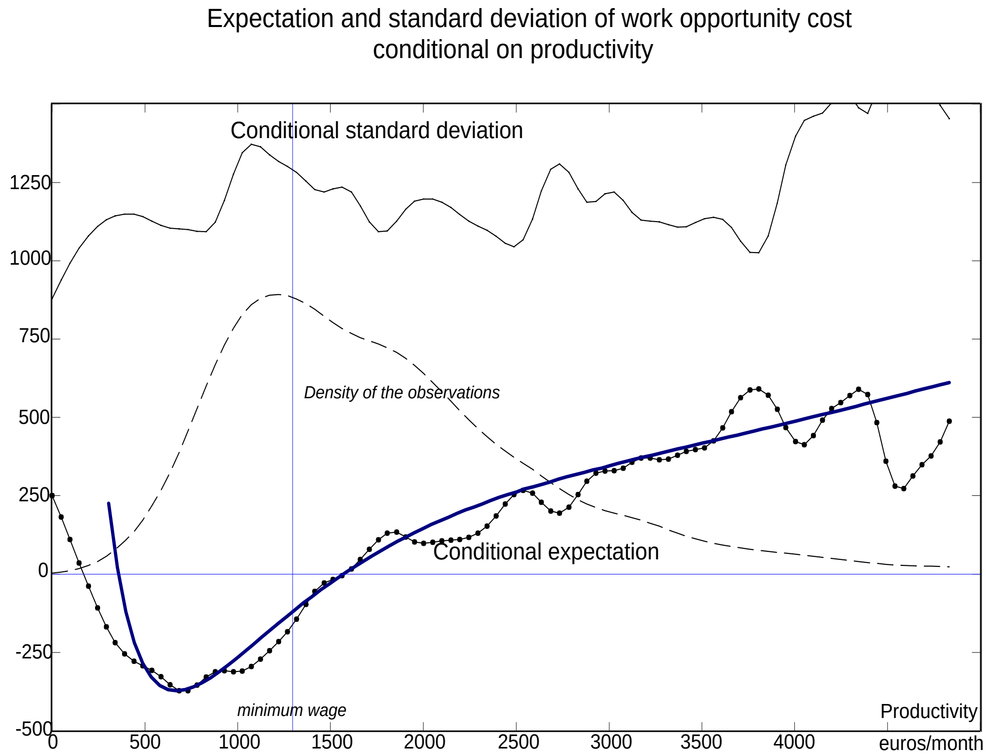

Figure 8: Non parametric regression of work opportunity costs on productivity

3.3 Computing the Laffer bound

According to (5) and (7), given a value of the observable variables (X, Z) (which include income when out of work $R(0) = r$; recall that the Laffer bound is defined conditionally upon income out of work) and of the unobserved heterogeneity η, the work opportunity cost of the typical agent is given by:

$$\Delta = Z\gamma + \rho \frac{\ln w - X\alpha}{\sigma_\varepsilon} + \sigma_\eta \eta. \tag{8}$$

If the tax schedule can be conditioned on the information available to the econometrician, for all wages above the minimum wage, the full information Laffer bound is solution of

$$\max_{d<w} (w - d) \Pr[\Delta \leq d] \; P_k(Y\beta),$$

using the assumption that keynesian unemployment is independent of η, or, if H denotes the c.d.f. of the variable η,

$$\max_{d<w} (w - d) H \left[\frac{1}{\sigma_\eta} \left(d - Z\gamma - \rho \frac{\ln w - X\alpha}{\sigma_\varepsilon} \right) \right]. \tag{9}$$

With this detailed information, which includes age at end of studies, experience, diplomas, marital status, number and ages of children,... the solution of (9) allows

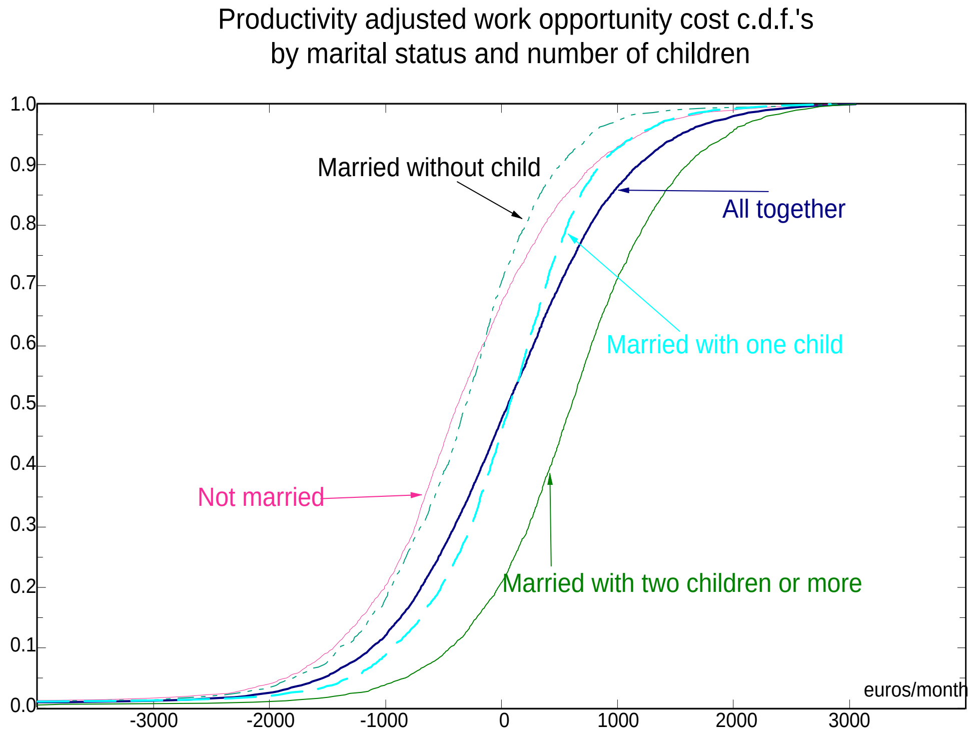

Figure 9: Distribution of work opportunity costs by marital status and number of children

the tax authorities to extract the maximum surplus from the public[15].

In my opinion, the above full information Laffer tax schedule is not of much practical interest. There are two reasons for this. First, parliaments or governments only allow tax schedule to be conditioned on relatively coarse information. Second, given the available data used here, to identify the distribution of η, I have assumed that it be identical across all the population. This is a strong assumption and I would not be confident to show results that deal with fine cases: I look only at broad categories, for which I need to get the distribution of work opportunity costs.

How does one get the distribution of work opportunity costs conditional on productivity for an arbitrary subset of the population? I have not found much guidance in the literature, but a possible way out is as follows. First, I simulate the residuals of the structural model, conditional on the observations, drawing (ε, ν, η) conditional on the observed employment status and on wage when employed[16]. This gives values of Δ associated with all the observations in the sample and therefore allows to compute the c.d.f. of Δ in any subpopulation[17].

To get the distribution of Δ conditional on w, I postulate a semi parametric relationship

$$\Delta(a) = \mu(w) + \sigma(w)\delta, \tag{10}$$

where $\mu(w)$ and $\sigma(w)$ are functions to be estimated and δ has a distribution that does not depend on w. Based on the previously simulated Δ's, I undertake both nonparametric and flexible functional form regressions of work opportunity costs and its square on productivity. Figure 8 shows that heteroscedasticity is not large and I have taken $\sigma(w)$ to be a constant. The mean of Δ appears to be varying

[15]The Laffer tax depends heavily on the shape of the function H, or equivalently on the participation elasticity $dH'(d)/H(d)$. The whole purpose of the estimation procedure was to impose *no* a priori constraint on H, and therefore to let the data determine the shape of the elasticities (even though, to save space, I did not report the non parametric results). The interested reader can check that a mixture of three logistics allows for many possible shapes.

[16]It would be easier, but probably less precise, to simulate the residuals unconditionally. I used the conditional residuals in part because I needed them in an alternative non parametric estimation method. They are computed with the help of the Gibbs sampling algorithm. For an unemployed person, given η and a value of ν smaller than $P_k(Y\beta)$, the difficulty is to draw a value of ε such that either $W(\varepsilon) < w_{\min}$ or

$$R[W(\varepsilon)] < R(0) + Z\gamma + \rho\varepsilon + \sigma_\eta \eta. \tag{a}$$

Let $\varepsilon_{\min}$ be the value of ε such that $W(\varepsilon)$ is equal to w_{min}. The half line $\varepsilon \geq \varepsilon_{\min}$ is divided into 50 equal probability intervals and $R[W(\varepsilon)]$ is tabulated at the median points of these intervals. Then ε is drawn from a conditional normal restricted to the union of $\varepsilon \leq \varepsilon_{\min}$ with the intervals such that (a) is satisfied at their median points.

[17]To conform with the model, since I am only interested in the computation of the Laffer taxes, I only keep in the following the observations that are not subject to keynesian unemployment, associated with women who would not work whatever the tax schedule.

with productivity, and the remainder of the paper uses the flexible functional form estimate of $\mu(w)$, represented with the bold solid line on the figure[18]. I then have a value of the residual $\sigma\delta$ of equation (10) attached to each observation. Let G be a kernel based estimation of its c.d.f.. The distribution of work opportunity costs, conditional on w, is a simple translation of G by $\mu(w)$.

Figure 9 shows a graph of G for the sample as a whole and a similar distribution function for some subcategories of the population. The c.d.f.'s appear to be smooth: while the estimated distribution of unobserved heterogeneity η has a mass point (Figure 7), the observed heterogeneity, stemming particularly from the wage of the spouse, smoothes the overall distribution. As expected, the distribution of work opportunity costs seems to be first order increasing with the number of children in the household.

The Laffer tax schedule is the solution of

$$\max_D (w - D)G(D - \mu(w)) = \max_d (w - \mu(w) - d)G(d).$$

It can be easily numerically computed[19]. Since the distribution of work opportunity costs depend on w, Assumption 3 is not satisfied, and the tax schedule cannot be expected a priori to have the properties listed in Theorem 4: it might not be increasing in w.

3.4 Is the French welfare state efficient?

A number of tools now are available to discuss the efficiency of the French tax benefit schedule. From Theorem 2, a necessary condition for efficiency is that taxes

[18]The regressors are w, its square, cube, square root and logarithm, as well as a constant term. While the estimates are very fragile below the minimum wage, it is worth noticing that $\mu(w)$ appears to be negative in this region. It seems that people of low productivities are eager to work, even if they lose money in the process. An implication is that the minimum wage legislation reduces both employment and potential tax revenues.

[19]There are (at least) two ways to do this computation. Brute force involves evaluating the function G^{-1} at, say, a thousand quantiles, and computing the maximum on a grid of points w of interest. Another possibility, given the smoothness of the distribution, is to use the first and second order conditions associated with the maximization. The first order condition is

$$w - \mu(w) = d + \frac{G(d)}{G'(d)},$$

and the second order condition is

$$-2[G'(d)]^2 + G(d)G''(d) < 0.$$

It is easy to get the derivatives of G by differentiating the kernel. The second order condition then yields a range of admissible values of d and the corresponding $w - \mu(w)$'s follow from the first order condition. This makes for smoother figures than the brute force technique and the figures shown below are drawn according to the latter method.

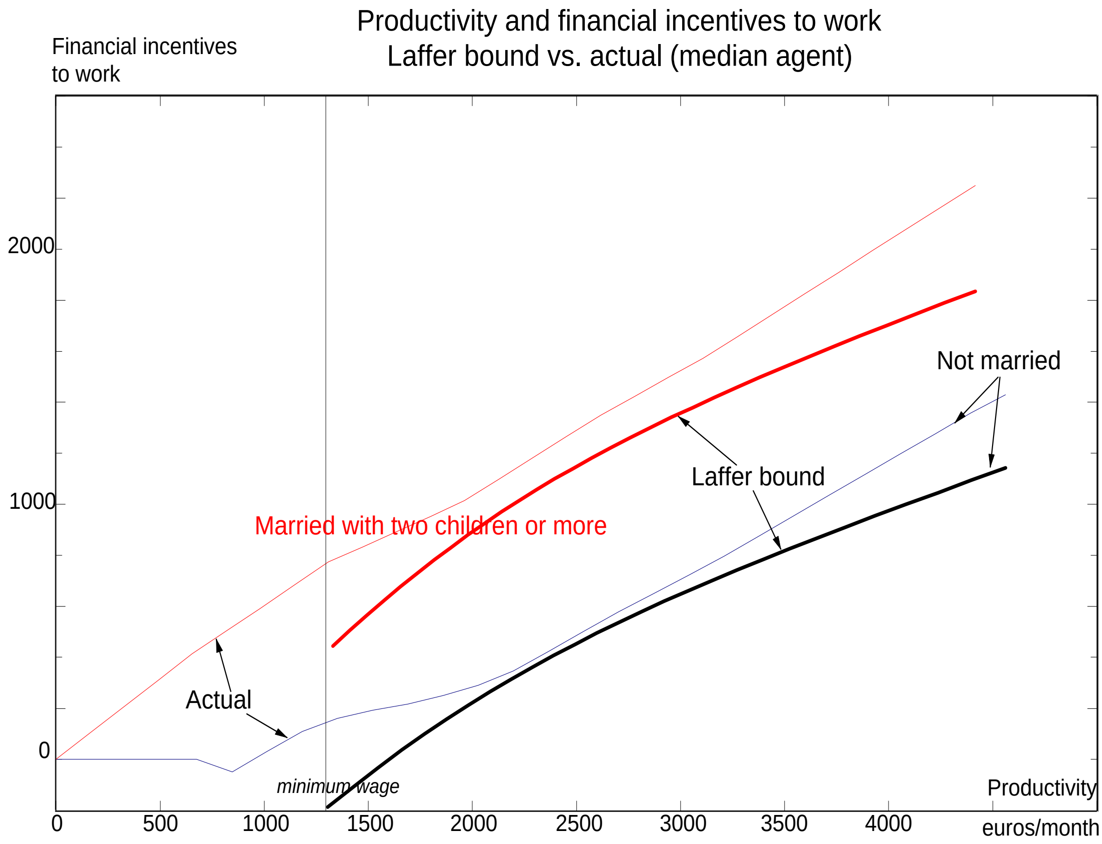

Figure 10: Laffer vs. actual incentives for the median agent

be lower than the Laffer bound. This condition is sufficient provided that (Assumption 1) work opportunity cost is non decreasing in the subsistence income, that financial incentives always are lower than productivity, and that (Assumption 2) there is no overtax. Furthermore, there is a procedure to compute the Laffer bound. Here, I shall concentrate on two broad segments of the population: single women, and women with two children or more.

The main assumption that needs consideration is the assumption that the work opportunity cost is non decreasing in the level of out of work income[20]. The model specifies a work opportunity cost that depends linearly on income out of work $R(0)$, the coefficient depending on the marital status and on the number of children of age smaller than 3, between 3 and 6, and over 6. For single women, the coefficient is 0.16, with standard error (0.05), significantly positive. For a woman with two children, one less than 3, the other between 3 and 6, the coefficient is -0.08 (0.05). It is not significantly different from zero, but the point estimate is negative. Women with young children use the opportunity of a higher income of their spouse to substitute taking care of the children at home with a job on the market. Any further young child, less than 6, adds -0.15, while older children

[20]Direct inspection of the tax schedule shows that $D(w;x)$ is smaller than w in the relevant region, for productivities larger than the cost of the minimum wage. Also the numerical computations indicate that Assumption 2 is satisfied in this region.

add 0.06 to the coefficient. To summarize, Assumption 1 seems to hold for single women, but is in doubt for mothers of two children or more.

Figure 10 shows the Laffer financial incentives to work (solid lines). They need to be compared to the actual incentives to work (dotted lines) provided by the French taxes and social transfers. I took a *median* representative agent of each of the categories, single and married with two children or more[21]. Several comments are in order:

1. It is comforting to see that the French system seems to be on the right side of the Laffer curve. However, the distance is small. It looks as if, through the interaction between the various government agencies and possible lack of intertemporal consistency (see Buchanan and Lee (1982)), the transfer schedule is close to maximizing government income on these two categories of women.

2. The results indicate that the system is second best efficient for single women. Since women with children have a work opportunity cost that may be increasing with their out of work income, the theory cannot assert that the tax benefit schedule is efficient for them. This should be the subject of further work.

3. Any second best optimum yields larger incentives than the Laffer bound. The Laffer bound itself corresponds to the preferred choice of a Rawlsian planner who would maximize government revenues to redistribute income to the poorer persons in society. The proximity of the actual French system to the Laffer bound suggests that the implicit social welfare criterion of French politicians, again concerning the women in the sample, is not far from Rawls.

4. The computation depends on the (assumed) information that the government is entitled to use in the design of the transfer scheme.

References

Beaudry, P., and C. Blackorby (1997): "Taxes and Employment Subsidies in Optimal Redistribution Programs," Discussion Paper 22, University of British Columbia.

Blundell, R., A. Duncan, J. McCrae, and C. Meghir (2000): "The Labor Market Impact of the Working Families' Tax Credit," *Fiscal Studies*, 21, 75–104.

[21]The values taken for each component of X and Z are the median values in the subsample under consideration.

BLUNDELL, R., A. DUNCAN, AND C. MEGHIR (1998): "Estimating Labor Supply Responses Using Tax Reforms," *Econometrica*, 66(4), 827–861.

BUCHANAN, J. M., AND D. R. LEE (1982): "Politics, Time and the Laffer Curve," *Journal of Political Economy*, 90(4), 816–819.

CANTO, V. A., D. H. JOINES, AND A. B. LAFFER (eds.) (1982): *Foundations of Supply-Side Economics, Theory and Evidence.* Academic Press, New York.

CARD, D., C. MICHALOPOULOS, AND P. K. ROBINS (1999): "When Financial Incentives Pay for Themselves: Early Findings from the Self-Sufficiency Project's Applicant Study," Discussion paper, SRDC, Social Research and Demonstration Corporation.

CHONÉ, P., AND G. LAROQUE (2001): "Optimal Incentives for Labor Force Participation," Discussion Paper 25, CREST-INSEE, forthcoming in the *Journal of Public Economics.*

DIAMOND, P. (1980): "Income Taxation with Fixed Hours of Work," *Journal of Public Economics*, 13, 101–110.

EISSA, N., AND J. LIEBMAN (1996): "Labor Supply Response to the Earned Income Tax Credit," *Quarterly Journal of Economics*, 111(2), 75–104.

HOROWITZ, J. (1998): *Semiparametric Methods in Econometrics*, no. 131 in Lecture Notes in Statistics. Springer.

LAROQUE, G., AND B. SALANIÉ (2000): "Une décomposition du non-emploi en France," *Economie et statistique*, 331, 47–66.

——— (2002): "Labour Market Institutions and Employment in France," *Journal of Applied Econometrics*, 17(1), 25–48.

MEYER, R., AND D. WISE (1983a): "Discontinuous Distributions and Missing Persons: The Minimum Wage and Unemployed Youth," *Econometrica*, 51, 1677–1698.

——— (1983b): "The Effect of the Minimum Wage on the Employment and Earnings of Youth," *Journal of Labor Economics*, 1, 66–100.

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

MOFFITT, R. A. (2002): "Welfare Programs and Labor Supply," in *Handbook of Public Economics, volume 4*, ed. by A. J. Auerbach, and M. S. Feldstein. North Holland.

Robins, P. K., and C. Michalopoulos (2001): "Using Financial Incentives to Encourage Welfare Recipients to Become Economically Self-Sufficient," *Federal Reserve Bank of New York Economic Policy Review*, 7(2), 105–123.

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

——— (2002): "Optimal Income Transfer Programs: Intensive versus Extensive Labor Supply Responses," *Quarterly Journal of Economics*, 117, 1039–1073.

Stiglitz, J. E. (1982): "Self-Selection and Pareto Efficient Taxation," *Journal of Public Economics*, 17, 213–240.