Income maintenance and labor force participation

Guy Laroque[1]

September 2003

[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 and at the meetings of the Econometric Society in Brisbane, Lahore, Los Angeles, Sao Paulo and Venice. Comments from Costas Azariadis, Richard Blundell, James Heckman and Rafael Repullo 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 aversions to work. 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 aversions 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.

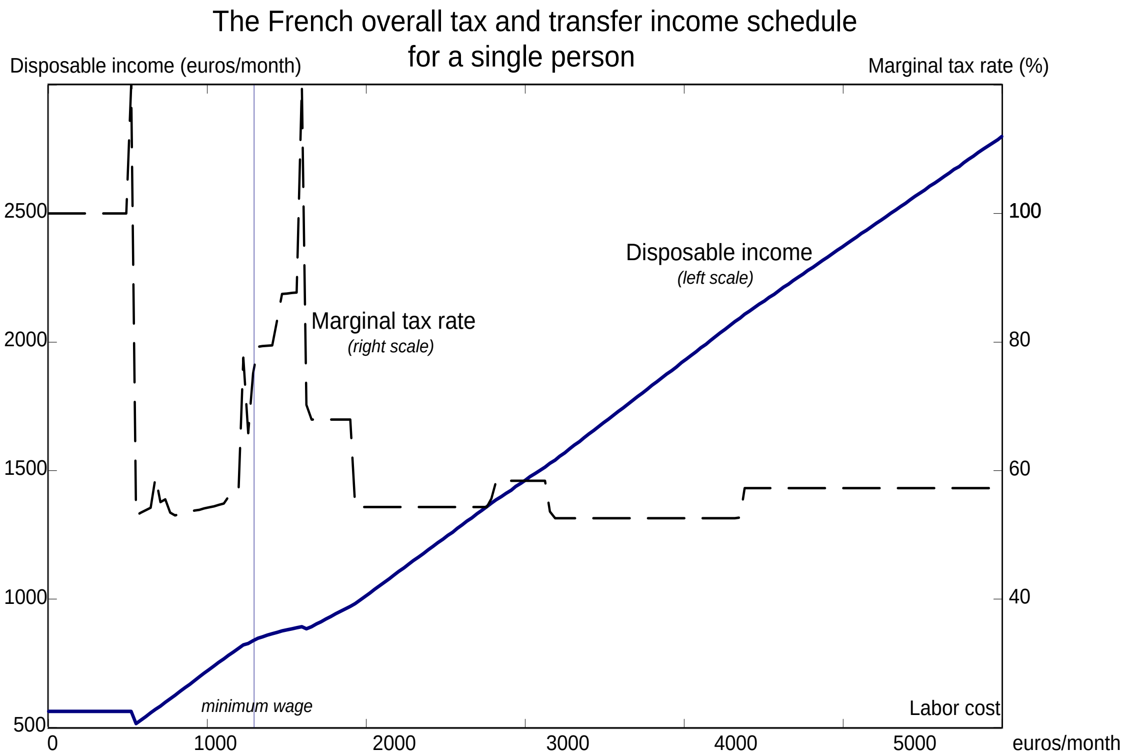


Figure 1: The French tax and transfer schedule in 1999

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 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 policy makers in the past thirty years around the world. 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[1], 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[2]. 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. 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 depends on the joint distribution of productivities and *work aversions* (or monetary *disutilities for work*) in the economy. A qualitative analysis shows that it involves positive work subsidies or *negative* marginal tax rates in a region where the c.d.f. is not log concave, for instance when it has some mass points. More generally, the shape of the Laffer bound directly derives from the distribution of work aversions: essentially any financial incentive scheme can be rationalized with an appropriate distribution of work aversions (Choné and Laroque (2001)).

The theoretical results show that tax schemes are little restricted by efficiency considerations and that reforms, say towards making work pay, should be ana-

[1]This is a big 'if'. The political forces indeed may lead to an inefficient allocation of resources.

[2]Also, the optimal tax program is quite difficult to solve. These difficulties have 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.

lyzed as political redistribution, provided that one stays on the right side of the Laffer bound. The economist should therefore provide a measure of this bound. The difficulty is to get an estimate of the joint distribution of work aversions and productivity. I rely on a labor supply model developed with Bernard Salanié on French data, which accounts for the minimum wage. It is estimated on a sample of women aged 25-49. Work aversion depends 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 implement a (flexible) estimation procedure.

Once the joint distribution of productivities and work aversions 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 a discrepancy 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. The first component of a is the (non negative) productivity of the agent. 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 aversion*[3]. Income, or consumption, is assumed to always be desirable. Consider an agent a who is indifferent between working with income C and not working with income c. Her work aversion, in either of these two situations, is the (possibly negative) difference $C - c$. The labor supply of the agent is fully characterized by the value of her work aversion, which can be measured alternatively as a function of income out of work, $\Delta(c; a)$, or of income at work, $\Gamma(C; a)$. By definition:

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

and

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

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

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

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

[3]An alternative terminology, suggested by Rafael Repullo, is *minimum inducement to work.*

One representation of this utility, which I shall use at times 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[4].

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. The assumption that Δ is nondecreasing in c is equivalent, in this setup, to leisure being a normal 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. 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 $\Delta(c, a) \geq 0$.

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

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

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 he 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

[4] When consumption is restricted to be positive, there are some difficulties in the correspondence between the two approaches, work aversions or utilities, at the lower boundary of the domain. For instance, if work aversion 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 aversions are not constrained, which accounts for agents with a negative work aversion, who would rather work on the market than stay at home with the same income.

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 market 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) 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 (second best) and the allocation has to be measurable with respect to the planner's information.

2.2 First best allocations

In this setup, it is easy to characterize the set of Pareto optimal allocations. Indeed, it is Pareto optimal that 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 this condition is simple. 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 (2) implies that the allocation is Pareto optimal. An allocation satisfies the feasibility constraint:

$$\int c(a)dF(a) = \int_{s(a)=1} w(a)dF(a).$$

Using the utility index defined above:

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

the constraint can be rewritten equivalently:

$$\int u[c(a), s(a), a]dF(a) = \int_{s(a)=1} w(a)dF(a) + \int_{s(a)=0} \Delta(c(a); a)dF(a). \tag{3}$$

Now consider a feasible allocation $(c(a), s(a))$ which satisfies the conditions of the Theorem and (3) , and suppose that there is another allocation $(c'(a), s'(a))$ which makes every agent at least as well off, and some strictly better off:

$$\int u[c'(a), s'(a), a]dF(a) > \int u[c(a), s(a), a]dF(a).$$

I claim that

$$\begin{aligned} \int_{s(a)=1} w(a)dF(a) + \int_{s(a)=0} \Delta(c(a); a)dF(a) \quad \geq \\ \int_{s'(a)=1} w(a)dF(a) + \int_{s'(a)=0} \Delta(c'(a); a)dF(a), \end{aligned}$$

so that $(c'(a), s'(a))$ is not feasible. Indeed, if $s'(a) = 1$ and $s(a) = 0$, the difference $w(a) - \Delta(c(a); a)$ is non positive by (2). When $s'(a) = 0$ and $s(a) = 1$, the difference $w(a) - \Delta(c'(a); a) = w(a) - \Gamma(c(a); a)$ is non negative by (2). ∎

At a Pareto optimal allocation, an agent works if her work aversion is strictly smaller than her productivity, does not work if her work aversion is strictly larger than her wage, and her status is indeterminate at equality. It follows that the market allocation is Pareto optimal. Note that work aversion 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 market equilibrium. A social objective that maximizes employment implicitly favors *laissez faire*.

Remark: The result sheds some light on an often asked policy issue. Consider a currently unemployed person, who collects benefits and contemplates taking a job. 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, reducing the wage cost, or to make work pay, increasing 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 provided that her work aversion is smaller than her productivity, at a potential extra cost, on top of the unemployment benefit, which can be as large as her work aversion.

Consider the particular case where work aversion does not depend on the productivity w, but only on x. Take a Pareto optimal allocation which gives equal utility to all agents of type x, independently of their productivities (a Rawlsian criterion would lead to such an allocation; also such an allocation would satisfy a fairness property for type x agents). Let $r(x)$ be the consumption of the agents x who do not work. Since the workers have the same utility level as the unemployed, they consume $R(x) = r(x) + \Delta(r(x); xa)$. Then Theorem 1 indicates that the workers are the agents with productivity w higher than the threshold $\Delta(r(x); x)$. Figure 2 represents the income collected by agent (w, x) as a function of w, for a fixed x. Agents with a productivity lower than $R(x) - r(x)$ do not work, while those with a larger productivity are employed. Every one with a work aversion larger than her productivity is kept out of the work force; all those whose work aversions are smaller than their productivities work. Pareto optimality implies full efficiency of the allocation of time between market and non market activities.

2.3 Second best allocations and tax subsidy schemes

I now turn to second best situations, where the distribution of characteristics in the economy is common knowledge, but the individual agent's aversion to 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

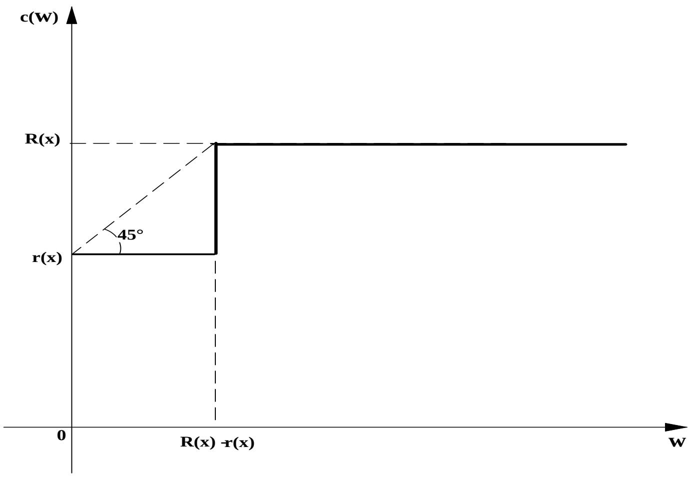


Figure 2: Fair first best allocations: work aversion is equal to productivity for the pivotal agent

schedule which describes the income of the agent according to her employment status and her productivity when she works.

When everyone has the same (constant) disutility Δ of work and the only information unknown to the government is the individual productivities, the first best result for fair allocations translates directly into an income schedule with a shape similar to that of Figure 2. After-tax income is equal to $r(x)$ for all before-tax wage income smaller than Δ, to $r(x) + \Delta$ when before-tax income is larger than Δ. This amounts to a marginal tax rate equal to $-\infty$ at the switching point, a large downward tax discontinuity. Such negative income taxes are not seen in practice. Quite the contrary, in France, apart from a temporary subsidy, there is a 100% marginal tax rate on earnings when one takes a job; in the USA, a similar feature was associated with the Aid to Families with Dependent Children program before the 1996 reform. It was partially mitigated by the Earned Income Tax Credit. The EITC can amount (at the maximum) to 40% of earnings, i.e. each dollar earned yields 1.4 dollar for the wage earner. After the welfare reform and the replacement of AFDC with TANF, there may exist some income schedules in some states with a zone of negative marginal tax rates. Still, this is far from the kind of discontinuities described above. But of course in practice work aversions are heterogeneous in the population and unobserved by the fiscal authorities: this fact is likely to smooth the shape of the optimal subsidy scheme, as will be seen below.

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 work aversion strictly lower than $D(w; x)$ get a rent, compared to the unemployment situation, which can be measured in monetary terms by $D(w, x) - \Delta(r(x); a)$. Therefore 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 aversions $\Delta(r; a)$ conditional on the agent productivity w

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

The probability that an agent of type x with productivity w works when she faces the schedule $(r(x), D(w; x))$ 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

$$\int [w - D(w; x)] G_{r,w,x}(D(w; x)) \mathrm{d}\tilde{F}(w, x) = \int r(x) \mathrm{d}F(a), \tag{4}$$

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

where $\tilde{F}$ is the joint distribution of productivities and observable characteristics in the population. The left hand side of (4) 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

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

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 non zero 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 revenue, 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

$$d_r(w;x) = \sup \underset{D, D \leq w}{\operatorname{argmax}} (w - D) G_{r(x),w,x}(D). \tag{5}$$

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:

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

or, equivalently, using the feasibility constraint (4):

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

The first term in 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.

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. ∎

Example: A Rawlsian planner maximizes the welfare of the least-advantaged persons in the economy. Since the planner knows the type x of the agents, his preferences are defined through some welfare measures or utility functions assigned to each type, say $v(c,s,x)$. In the second best environment, since every type x agent can get $r(x)$ if she does not work, the Rawlsian planner typically maximizes $\inf_x v(r(x), 0, x)$. Under the feasibility constraint (4), it is easy to see (see Choné and Laroque (2001) for a formal argument) that this implies maximizing government revenue so as to have the largest possible transfers, that is making incentives equal to their lower bound of Theorem 2.

2.5 Characterization of second best optima

The condition of Theorem 2 is close to being a sufficient condition for a feasible tax schedule to be optimal. 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 be the case that there are larger incentives which yield the same government revenue, and therefore make all the workers better off, while keeping the existing transfers feasible. The following assumption, which says that government revenue decreases with the level of incentives above the Laffer bound, rules out this possibility:

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

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

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

Theorem 3 *Under Assumptions 1 and 2, any feasible tax schedule with incentives $D(w; x)$ larger than the Laffer bound, $D(w; x) \leq w$, supports a second best optimal allocation.*

Proof: I give the argument for an economy with a single type x, leaving the generalization to the interested reader. The feasibility constraint (4) is:

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

Consider two tax schemes $(r, D(w))$ and $(r', D'(w))$ with incentives above the Laffer bound, and 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 that government revenue decreases with D.

Suppose now that $r' > r$. By Assumption 1, $G_{r,w}(D)$ is non increasing in r. Therefore $(w - D(w))G_{r',w}(D(w)) \leq (w - D(w))G_{r,w}(D(w))$, since $D(w)$ is smaller than w by assumption. Therefore:

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

For the prime allocation to satisfy (4), since $(w - D)G_{r',w}(D)$ decreases in D, $D'(w)$ must be smaller than $D(w)$ on a non negligible set of w's, say W. On W, $G_{r',w}(D'(w)) \leq G_{r,w}(D(w))$. It follows that:

$$r \geq \int_{\mathbb{R}_+ \setminus W} [w - D(w)]G_{r',w}(D(w))\mathrm{d}\tilde{F}(w) + \int_W [w - D(w)]G_{r',w}(D'(w))\mathrm{d}\tilde{F}(w).$$

Furthermore, for the working agents to be as well off as in the reference allocation, one must have $r' + D'(w) \geq r + D(w)$. Now rewrite the preceding inequality as:

$$\begin{aligned} & r\left(1 - \int_W G_{r',w}(D'(w))\mathrm{d}\tilde{F}(w)\right) \\ \geq \quad & \int_{\mathbb{R}_+ \setminus W} [w - D(w)]G_{r',w}(D(w))\mathrm{d}\tilde{F}(w) + \int_W [w - D(w) - r]G_{r',w}(D'(w))\mathrm{d}\tilde{F}(w) \\ \geq \quad & \int_{\mathbb{R}_+ \setminus W} [w - D'(w)]G_{r',w}(D'(w))\mathrm{d}\tilde{F}(w) + \int_W [w - D'(w) - r']G_{r',w}(D'(w))\mathrm{d}\tilde{F}(w), \end{aligned} \tag{6}$$

where the substitution of the first term of the left hand side comes from the fact that government revenue decreases with D, and the second from the inequality $r'+D'(w) \geq r+D(w)$ on W. Now, if the employment rate in the prime allocation $\int_W G_{r',w}(D'(w))\mathrm{d}\tilde{F}(w)$ is strictly smaller than 1,

$$r'\left(1-\int_W G_{r',w}(D'(w))\mathrm{d}\tilde{F}(w)\right) > r\left(1-\int_W G_{r',w}(D'(w))\mathrm{d}\tilde{F}(w)\right),$$

and using (6),

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

which implies that $(r', D'(w))$ is not feasible, the desired contradiction.

If $\int_W G_{r',w}(D'(w))\mathrm{d}\tilde{F}(w)$ is equal to 1, W is the full set, everyone works both at the reference and at the prime allocation. Since the aggregate resources are given, equal to $\int w(a)dF(a)$, the prime allocation cannot strictly Pareto dominate the other one.

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 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 fact that government revenue decreases with D, since $\bar{D}(w) \geq D(w)$, with some strictly positive inequalities. ■

Remark: Assumption 2 has an intuitive content, but is restrictive. It is satisfied whenever the distribution of work aversions $G(D)$ is log-concave (see Theorem 5 for further properties of the Laffer tax schedule when $G(D)$ is log-concave). Note that a simple adaptation of the proof shows that

> *under Assumption 1 any feasible tax schedule with incentives $D(w;x) \leq w(a)$ and such that there is no $D'(w;x) > D(w;x)$ yielding as large*

government revenue, $\int (w - D'(w;x))G_{r,w,x}(D'(w;x)dF(a) \geq \int (w - D(w;x))G_{r,w,x}(D(w;x)dF(a)$ *supports a second best optimal allocation.*

Remark: In a first best optimum, nobody works when her productivity is smaller than her work aversion. Theorem 3 only studies situations where the incentives to work $D(w;x)$ are smaller than w, so that nobody with a work aversion larger than her productivity will choose to work. However, contrary to the first best, there may exist 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.

The second best allocations that satisfy the assumptions of Theorem 3 can be described as follows. Take any non negative function on X, say $\lambda(x)$, which is integrable with sum equal to one, and describes the relative weights that the planner puts on the welfare of the unemployed agents of type x. Let ρ be a non negative number, and look for the value of ρ such the allocation which gives the income guarantee $\rho\lambda(x)$ to type x agents when the government maximizes its revenue, i.e. announces incentives $d_{\rho\lambda(x)}(w;x)$, is feasible:

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

The left hand side is continuous (see Choné and Laroque (2001), Lemma C.2) and non decreasing in ρ under Assumption 1. It is non negative when ρ is equal to zero. For each $\lambda(.)$, the above equation therefore has a unique root, say ρ_λ. By construction, any income guarantee larger than $\rho_\lambda\lambda(x)$ cannot be financed, since it would yield a government revenue less than ρ_λ. In the space of revenues $r(x)$, x in X, when the direction λ varies, the surface generated by the point at a distance ρ_λ from the origin describes the frontier of maximal feasible income guarantees. Any income guarantee $r(x)$ below the frontier, i.e. such that there exists $\lambda(.)$ with $0 \leq r(x) \leq \rho_\lambda\lambda(x)$ for all x, is associated with (many) second best allocations. There are typically a continuum of feasible tax schemes associated with $r(x)$, depending on the distribution of incentives among the workers of different productivities and types. Recall that under Assumption 2 government income is non increasing in $D(w;x)$. The only constraints[6] on these tax schemes are that they satisfy the Laffer bound, $D(w;x) \geq d_r(w;x)$, and that they finance the income guarantee, $\int [w - D(w;x)]G_{r,w,x}(D(w;x))\mathrm{d}\tilde{F}(w,x) = \int r(x)dF(a)$. The further 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,

[6]In case the agents can lie on their productivities, and pretend to be less productive than they are, truthful revelation would require $D(w;x)$ to be nondecreasing in w.

efficiency imposes essentially no constraint on the local shape of the tax scheme: it allows, for instance, for negative marginal tax rates.

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 aversions does not depend on r, nor on productivity w. Suppose that the work aversions and productivities both take their values in $[1, +\infty)$ with c.d.f. respectively

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

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

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

From Theorem 3, for r smaller than ρ, 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

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

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: 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[7]

The only requirement on the second best tax schedules is to be less strict than the Laffer tax. It therefore is of interest to study the properties of the Laffer tax, i.e.

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

of the government revenue maximizing problem. It turns out that the problem has an interesting structure which is most transparent when the distribution of work aversions is independent of the productivity of the agents:

Assumption 3 *The conditional distribution of work aversions $G_{r,w,x}(.)$ 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 aversions.

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 subgradient of $K(w)$, whose slope cannot thus exceed 1. Convexity implies that the subgradient is nondecreasing, which implies that $G_{r,x}(d_r(w; x))$ is nondecreasing in w, and $d_r(w; x)$ as well. ∎

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 3, 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,

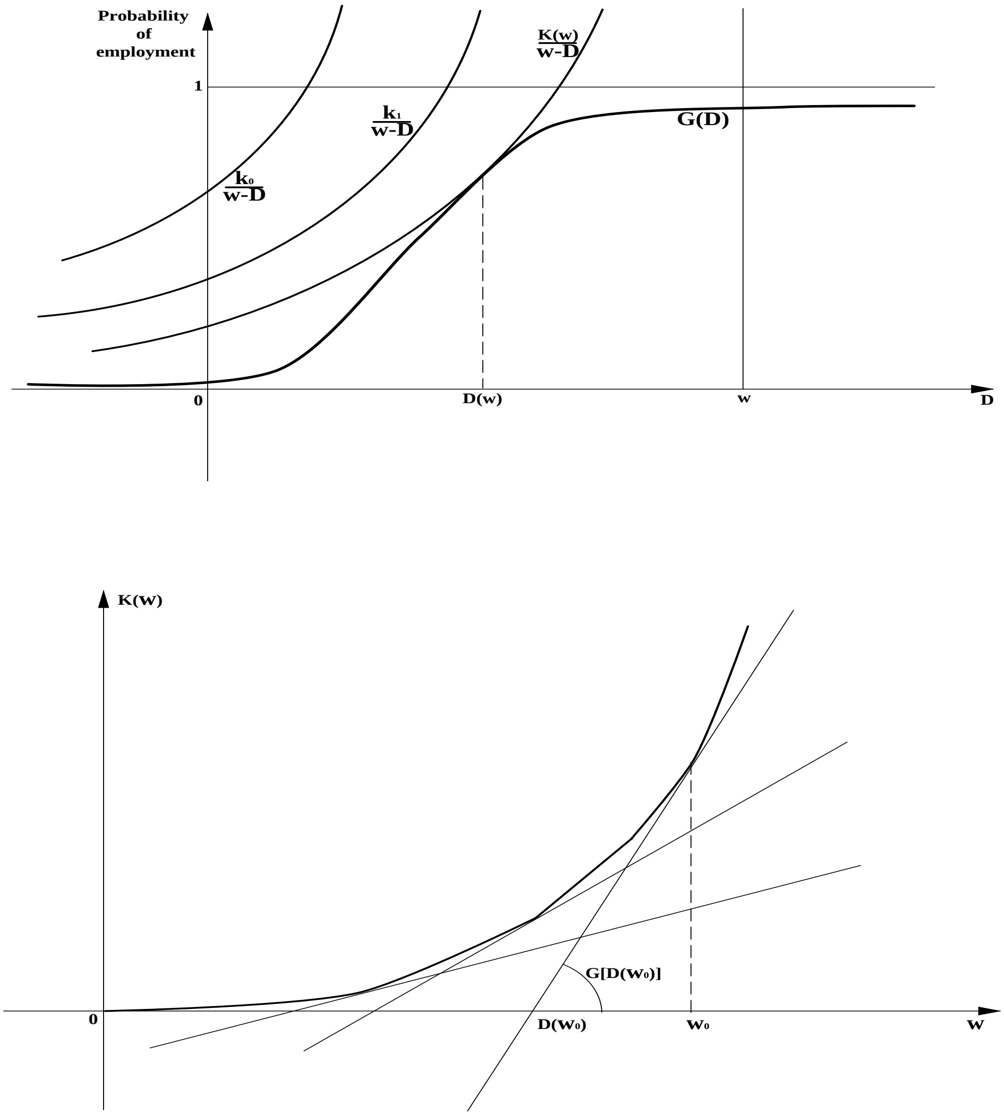


Figure 3: The optimization program

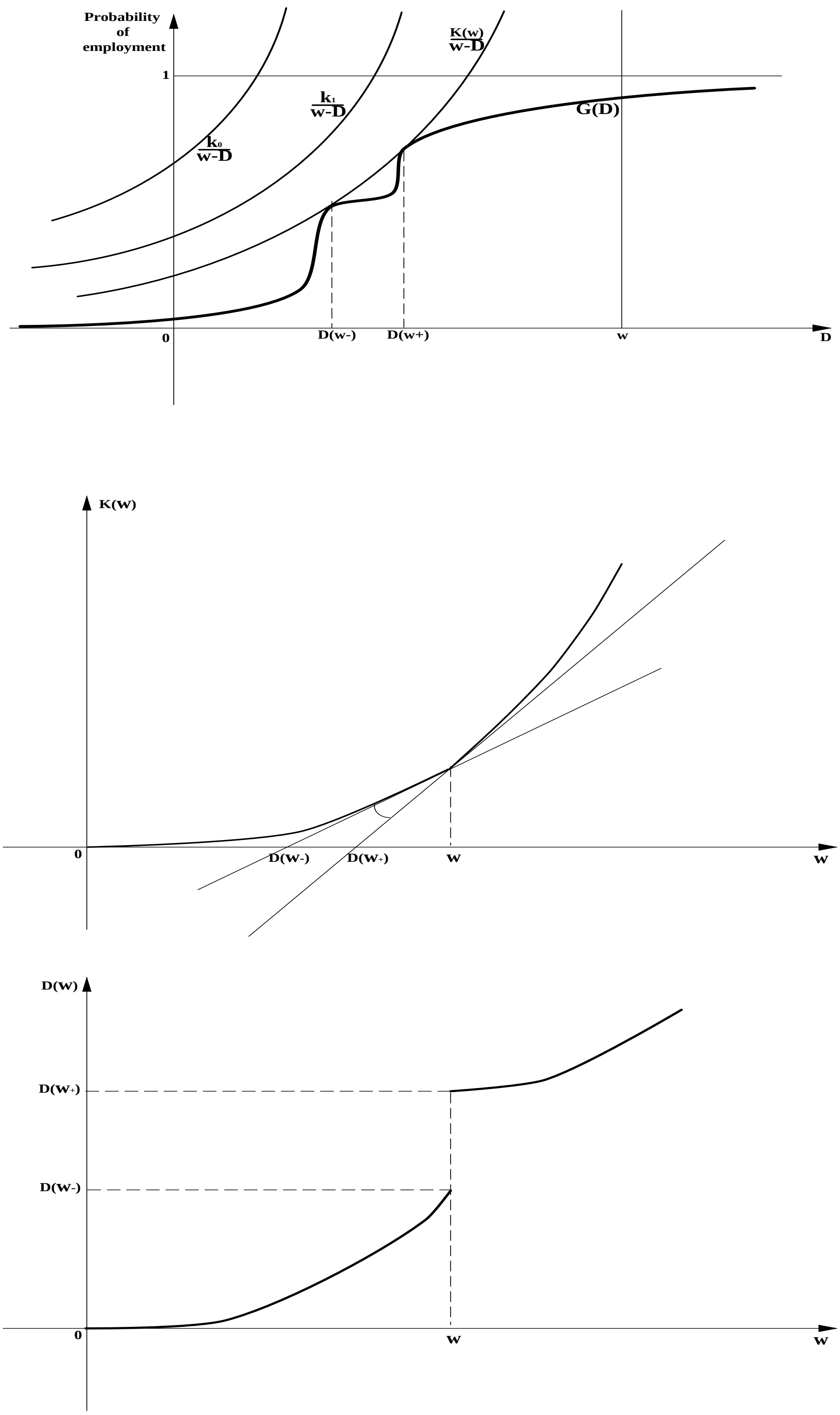


Figure 4: Discontinuity of the tax scheme

so that both $d_r(w)$ and $K_r(w)$ increase. It is also of interest to locate the market allocation on the graph. For r equal to zero, all the agents with a work aversion 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, needs not be well behaved. However, the overall optimization is simple, as shown on the bottom panel of Figure 3 drawn in the plan $(w, K(w))$. 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 4 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 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 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 indeed optimal for some distribution of work aversion (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 aversion close to d (d being known to the planner). Recall that work aversion 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 work aversion is very concentrated around d, the second best solution 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 aversion 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 (5) can be rewritten $\max_{D \le w} \ln(w-D)+H(D)$, 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[8]

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

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 in itself is of little guidance as to the shape of the optimal income support schedules, which are determined by the distributions of work aversions in the population. To put the theory to practical use, one has to postulate and estimate a labor supply model to derive the distributions we are looking for.

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[9]. 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 :

$$\ln w = X\alpha + \sigma_\varepsilon \varepsilon, \tag{7}$$

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

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

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

[9]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. Also the civil servants who have tenure are excluded from the sample under study.

2. She is not subject to frictional or keynesian unemployment

$$\nu \leq P_k(Y\beta), \tag{8}$$

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.

$$R(w) \geq R(0) + Z\gamma + \rho\varepsilon + \sigma_\eta \eta. \tag{9}$$

Here $R(.)$ is a known highly non linear function: $R(w)$ is the net after tax and subsidies income 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).

The unobserved heterogeneity is described by the triple (ε, ν, η). The model is estimated by maximum likelihood under the assumption that the three random terms are independently distributed. It is assumed that ε is distributed as a standard normal, ν as a uniform on the interval $[0,1]$, and η as a logistic. Under this parametric assumption, it is easy to recover the distribution of the work aversion Δ, equal to $Z\gamma + \rho\varepsilon + \sigma_\eta \eta$, which we are interested in (see below).

3.1 Semiparametric identification and estimation

The above model depends heavily on the assumed distributions of heterogeneity. It is of interest to see whether the data allow to identify these distributions, while keeping with the chosen exogenous variables and the functional forms used to describe their influence. The model is complicated and I shall follow a piecemeal approach, using intuitions from results in the literature obtained in simpler setups for models with a single or two equations, without formal proofs.

3.1.1 Identification

1. *Minimum wage.* The cost of the minimum wage relative to the distribution of observed wages appears to be high in France, by comparison with other developed countries. This potentially creates difficulties to identify the distribution of heterogeneity in the wage equation. Indeed, if one does not have some observed exogenous variables determining productivity (here mainly the diploma), there is no hope to know the distribution of potential wages *below* the minimum wage (Meyer and Wise (1983a) or Meyer and Wise (1983b)). Semiparametric identification here relies on the assumption that the distribution of heterogeneity does not depend on the diploma, 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, we take a very simple specification for this type of unemployment, which only depends on age, and does not include the exogenous variables which determine productivity (age at end of school, diplomas). Then, provided long studies and high diplomas characterize a set of persons who want 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 I am looking for.

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, initially assumed to be logistic, 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 (9) seems to be identified under location and scale normalizations, since there is a continuous variable among the Z's, the income of the spouse, which is an implicit argument of $R(0)$.

The above arguments discuss in turn the semiparametric 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. It may be of some comfort to know that, as far as ε is concerned, lognormality is not rejected by the data holding the distributions of ν and η at the maintained hypothesis (see Laroque and Salanié (2002) for a test against a mixture of lognormals).

3.1.2 Estimation of the distribution of η

According to the above argument, one should be able to estimate the distribution of η. At the very least, it seems desirable to check whether the parametric assumption of a logistic under which the first estimation has been carried out is acceptable. This is potentially of practical importance, since in the simulations below, the standard error of the estimated work aversion is of the order of 1150 euros, while the standard error of the unobserved heterogeneity term amounts to 850 euros. I have explored some of the various possible directions that seem open at this stage. It turns out that the unknown distribution of η only enters the likelihood function through its cumulative distribution function. This makes it particularly easy to use an adaptive estimation technique[10]: start with a maximum likelihood estimation of the parametric model; then estimate the c.d.f. of η, given the parameter values from the previous step (ignoring the assumed distribution of η) through a suitable semiparametric procedure; iterate the maximum likelihood estimation with the computed c.d.f., numerically interpolated, instead of the initial logistic, keeping fixed the location and scale parameters in equation (9) at their first step values. The crucial element of the procedure of course is the estimation of the unknown c.d.f.. I have looked at several possibilities.

1. The first idea that comes to mind is to use equation (9) to directly estimate the distribution of $R[W(\varepsilon)] - \rho\varepsilon - \sigma_\eta \eta$, where according to (7) $W(\varepsilon) = \exp(X\alpha + \sigma_\varepsilon \varepsilon)$. Indeed, if e is the indicator variable for employment, equal to 1 when the woman is employed and to zero otherwise, the following equality holds:

   $$E\left\{\frac{e}{P_k(Y\beta)}|Y\right\} = \Pr\{[R[W(\varepsilon)] - \rho\varepsilon - \sigma_\eta\eta \geq R(0) + Z\gamma]|W(\varepsilon) \geq w_{\min}, Y\}.$$

   A non parametric regression of $e/P_k(Y\beta)$ on $R(0) + Z\gamma$ therefore yields the cumulative distribution of $\rho\varepsilon + \sigma_\eta\eta - R[W(\varepsilon)]$, conditional on $W(\varepsilon)$ being larger than the cost of the minimum wage, and on the variables Y (diploma, age). To obtain the distribution of η, one faces a deconvolution problem, since the distribution of $R[W(\varepsilon)] - \rho\varepsilon$, conditional on $W(\varepsilon)$ larger than $w_{\min}$ and on Y, is known. Note that the function $R[W(\varepsilon)]$ is highly nonlinear and depends on a number of exogenous variables, such as the household composition and the spouse income. Unfortunately, this makes the deconvolution problem not tractable in practice.

2. To handle the difficulties associated with the function $R[.]$, a more promising road is to simulate the residuals of the structural model, conditional on the observations. For each observation, it is relatively easy to draw the residuals (ε, ν, η) in their joint distribution conditional on the observed employment status, and on wage when employed[11] [12]. I then use simulated residuals to

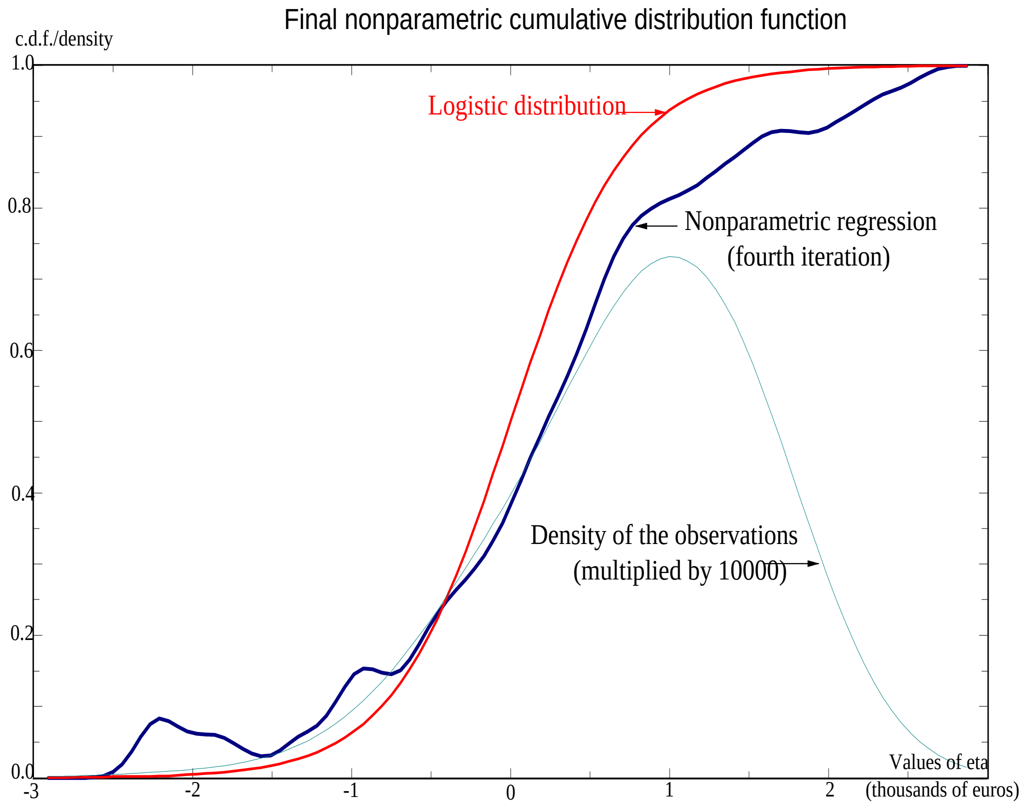
[10] A similar technique might be used more generally, in the estimation of structural models, when the distributions of the random terms are semiparametrically identified. However the implementation would typically involve more complicated functions of the unknown distributions than the mere c.d.f..

[11] This was done through 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

$$R[W(\varepsilon)] < R(0) + Z\gamma + \rho\varepsilon + \sigma_\eta\eta. \qquad (*)$$

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 $(*)$ is satisfied at their median points.

[12] In the tradition of the generalized residuals literature (Gouriéroux, Monfort, Renault, and



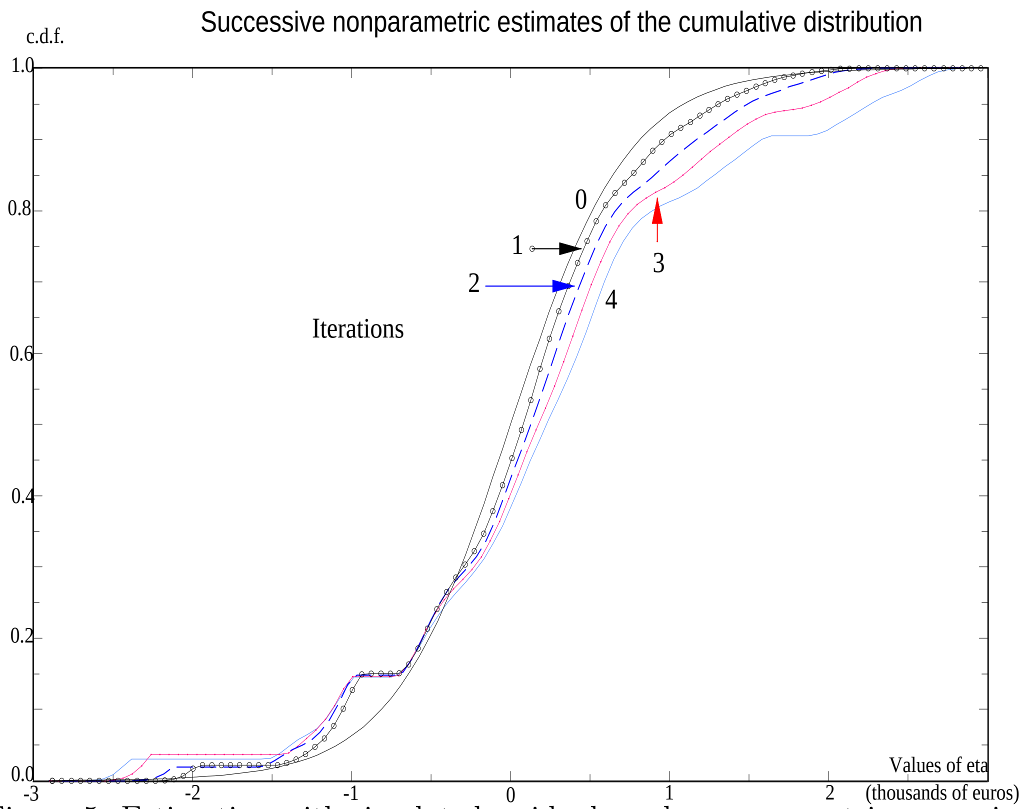


Figure 5: Estimation with simulated residuals and non parametric regression

implement a nonparametric regression of the observed employment status on the simulated value of

$$\frac{R(w) - R(0) - Z\gamma - \rho\varepsilon}{\sigma_\eta},$$

on the subset of observations such that, when a woman is unemployed, her simulated productivity is larger than $w_{\min}$ and she is not subject to frictional unemployment ($\nu \leq P_k(Y\beta)$). Two difficulties pop out when implementing this estimation strategy. First, the nonparametric regression is not constrained to be nondecreasing, while the c.d.f. has to be. Since the decreasing parts of the curve are located at the edges and are minor, I just fix the problem by taking the largest nondecreasing function that is everywhere smaller than the nonparametric regression[13]. Second, similarly, the nonparametric regression lacks precision in the range of values where there are few observations, here at the tails of the distribution. In the present situation, this eventually makes it impossible to pursue the algorithm. The upper tail of the distribution gets more and more weight as shown on the lower panel of Figure 5, but this comes from very few points since the density of the observations is of course low in this region (upper panel of Figure 5). The values of the (semiparametric) loglikelihood function at the maximum along the iterations are: -7902.12 (starting point), -7901.00 (first iteration), -7904.72 (second), -7905.60 (third). I have proceeded using the results of the last iteration, while checking how much difference this makes from the starting point.

3. A simpler procedure is to allow for some flexible functional form for the c.d.f. of η. Starting from the initial specification, I fix the location parameter (constant term) and the scale parameter σ_η and allow the random variable

Trognon (1987)), one then can build the c.d.f. of the simulated η, and, if different from the logistic, take it as a new distribution to estimate new parameters and iterate the procedure. However this estimation technique seems to be under too much influence from the starting point. To check its efficiency, I simulated a probit model with data generating process :

$$e = \mathbb{1}_{X+\eta>0},$$

where X has a centered normal distribution with standard error equal to 3, and η is an independent standard centered normal. I then computed the simulated residuals of this model, knowing X (there is no parameter to estimate), under the assumption that the distribution of η is logistic. The distribution of the simulated residuals, appears to be very close to the logistic, and thus quite far from the true (normal) distribution. However a nonparametric regression of e on X, $E(e|X) = 1 - \Pr(\eta < -X)$, works wonders, which motivates the route followed in the main text.

[13] This explains the difference between the nonparametric regression at the fourth iteration in the upper panel of Figure 5, which has some decreasing pieces, and the corresponding nondecreasing c.d.f. in the lower panel.

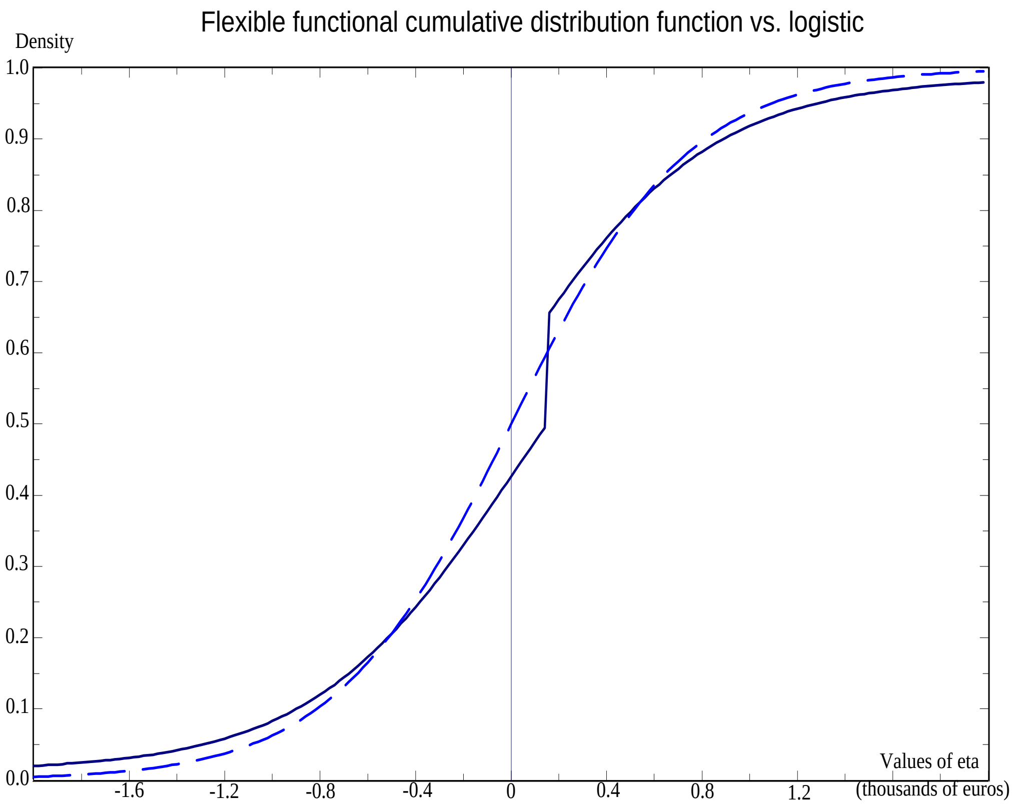


Figure 6: Flexible functional form

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

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

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

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 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. In the remainder of the paper, I shall retain this specification, which yields the best fit and the larger likelihood.

3.2 Is the French welfare state efficient?

Following Theorem 3, I am going to check that the actual tax-benefit schedule is on the right side of the Laffer bound[14]. Therefore the task is to compute the Laffer bound. The simulation of the model, conditional on the observations, yields a measure of the work aversion of each individual in the sample, which depends both on the observed characteristics of the household (presence or not of a spouse, income of the spouse, number of children, etc.) and on the simulated unobserved heterogeneity. Figure 7 shows a kernel based estimation of the cumulative distribution function of work aversions, for the sample as a whole and 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, the observed heterogeneity, stemming particularly from the wage of the spouse, smoothes the overall distribution. As expected, the distribution of work aversions seems to be first order increasing with the number of children in the household.

[14]The estimation yields a work aversion which is increasing in $R(0)$, so that Assumption 1 is satisfied. 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.

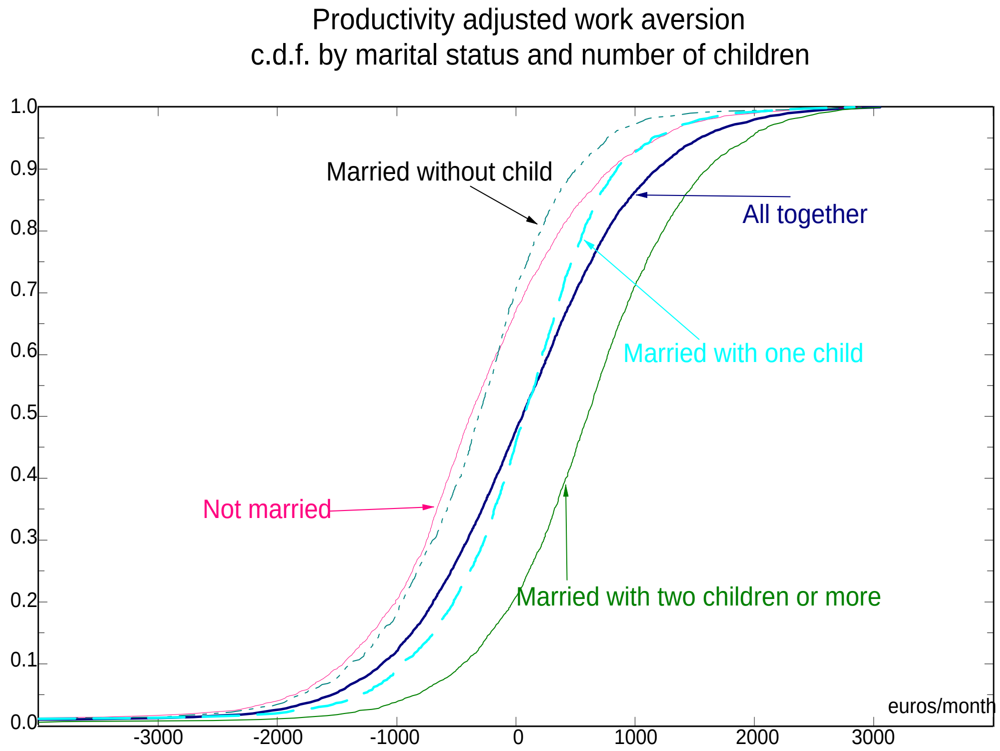


Figure 7: Distribution of work aversions by marital status and number of children

According to both the theoretical and empirical specifications, these c.d.f.'s depend both on the disposable income when not working and on the productivity of the agents. In all the computations below, I shall take income when out of work, the 'subsistence' income, as given. The government provides different maintenance incomes to different households, depending for instance on family composition and on the ages of the children. Also, income when out of work depends on the spouse's income.

The distribution of work aversions depends on productivity. Indeed, from equations (7) and (9),

$$\Delta(a) = Z\gamma + \rho \frac{\ln w - X\alpha}{\sigma_\varepsilon} + \sigma_\eta \eta. \tag{10}$$

If the government would design a different tax schedule for each value of (X, Z), one could just use the above formula, together with the estimated distribution of η, to compute the Laffer scheme. In practice, equity considerations, sometimes imbedded in the constitution (tax, for instance, cannot vary with the sex of the taxpayer) prevent the government from discriminating in such detail, and I look at broader categories. Here, I shall concentrate on two cases: single women, and women with two children or more.

The final difficulty comes from the fact that productivity w is correlated with the (unobservable to the government) exogenous variables X and Z. To tackle the

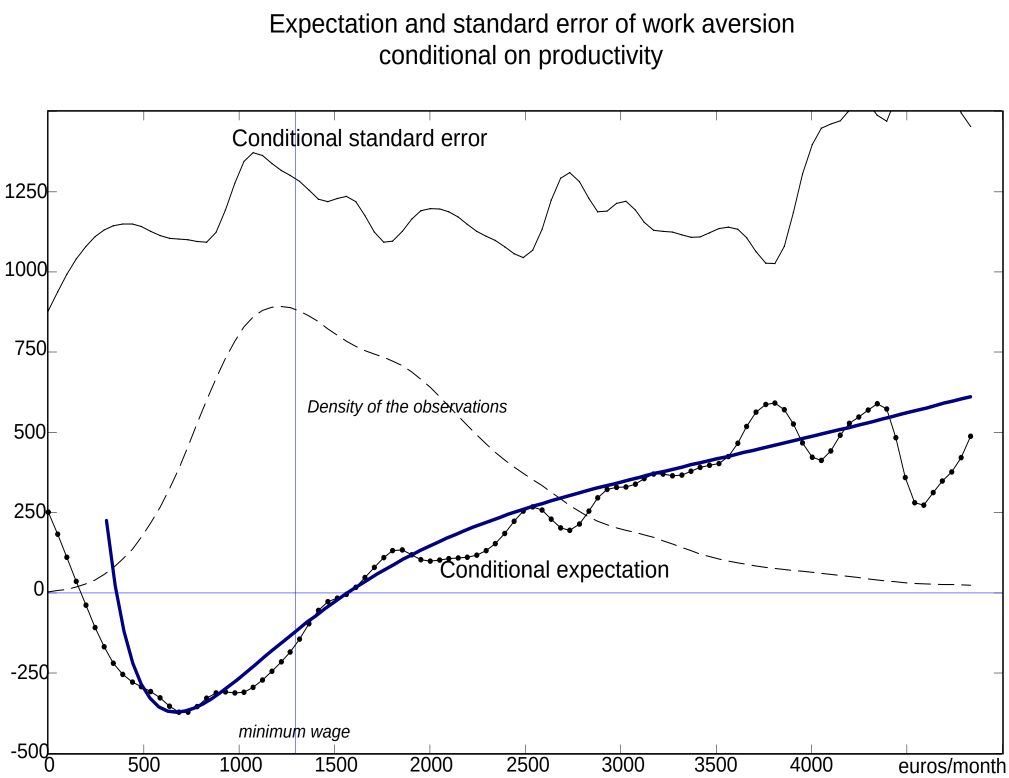


Figure 8: Non parametric regression of work aversion on productivity

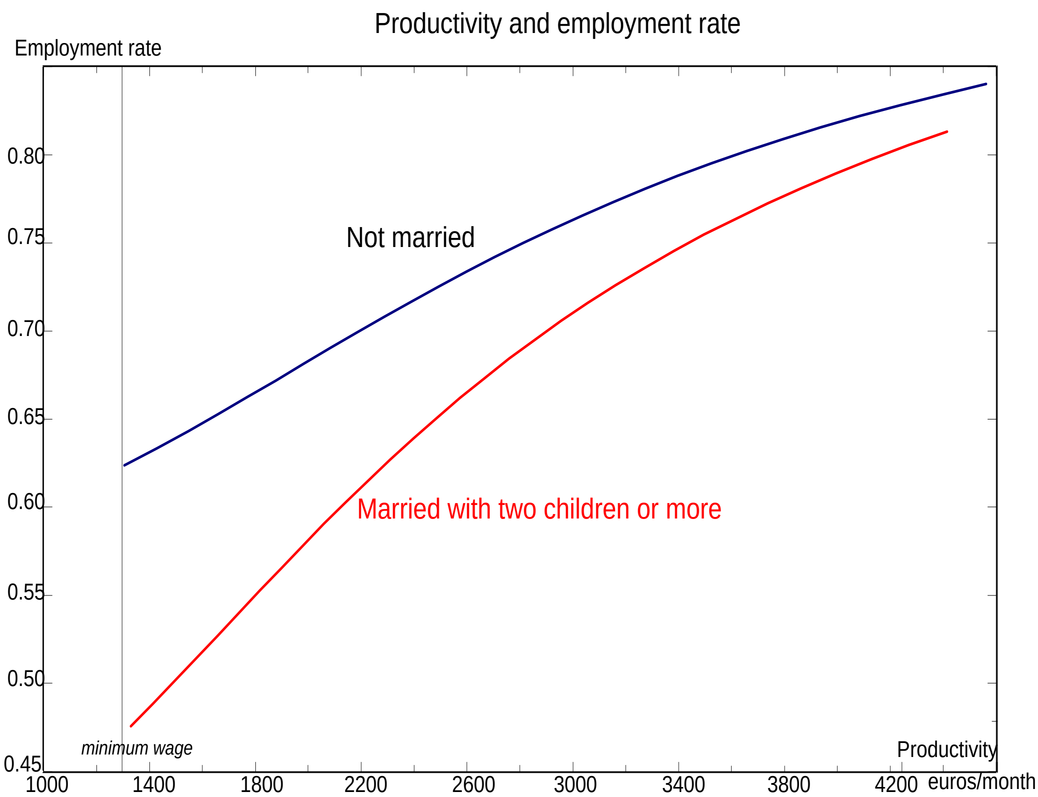


Figure 9: Laffer bound for single women and for married women with two children or more

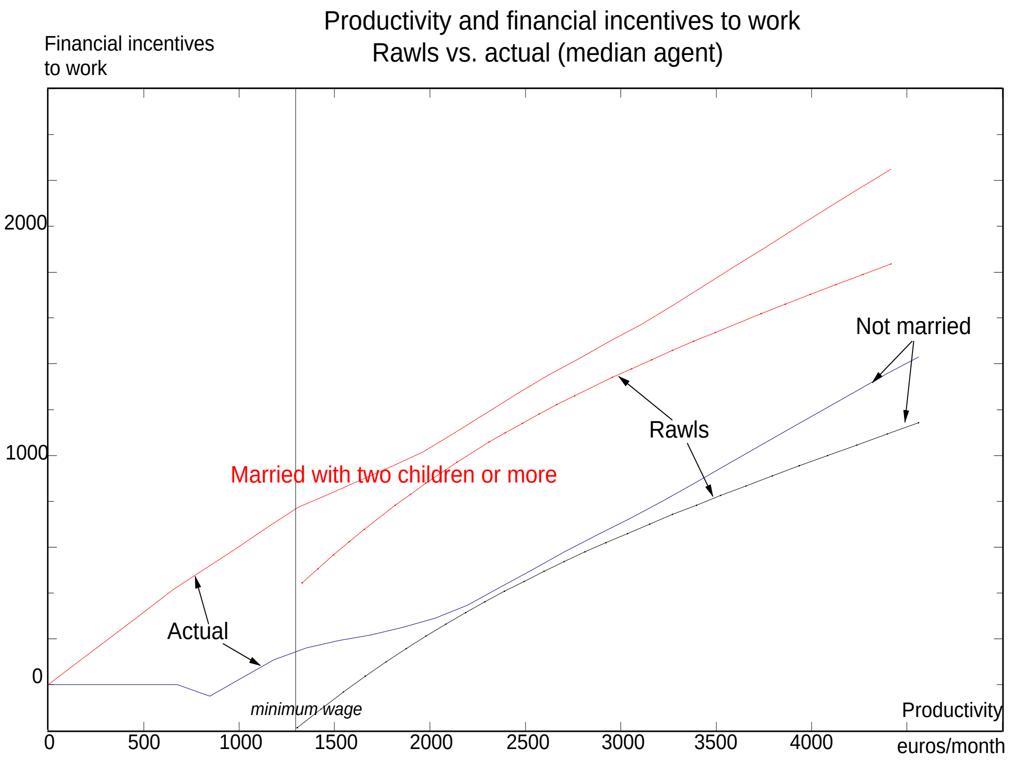


Figure 10: Laffer vs. actual incentives for the median agent

issue in a way that does not specify an a priori functional form as in equation (10) while allowing for this correlation, I postulate that

$$\Delta(a) = \mu(w) + \sigma(w)\delta, \tag{11}$$

where $\mu(w)$ and $\sigma(w)$ are functions to be estimated and δ has a distribution that does not depend on w. Let G_w be the c.d.f. of Δ and G the c.d.f. of δ. Then $G_w(D) = G[(D - \mu(w))/\sigma(w)]$. The revenue maximizing problem can be rewritten, letting $d = (D - \mu(w))/\sigma(w)$,

$$K(w) = \max_D (w - D)G_w(D) = \max_d (w - \mu(w) - \sigma(w)d)G(d) = \sigma(w)\kappa[\omega(w)],$$

where

$$\kappa[\omega] = \max_d (\omega - d)G(d) \quad \text{and} \quad \omega(w) = \frac{w - \mu(w)}{\sigma(w)}.$$

The function $\kappa[\omega]$ has all of the theoretical properties shown for $K(w)$ when the distribution does not depend of w in the qualitative analysis of 2.6: it is convex, increasing, with slope smaller than one, and the optimal d is nondecreasing in ω and everywhere smaller than ω. It can be easily numerically computed, which yields the Laffer tax schedule[15]. By construction, the optimal $D(w) = d[(w - \mu(w))/\sigma(w)] + \mu(w)$ is also smaller than w, but this is the only property that

[15]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

is preserved in general, everything else depending on the shape of the functions $\mu(w)$ and $\sigma(w)$.

In order to estimate the unknown functions $\mu(w)$ and $\sigma(w)$ of equation 11, I take the values of the disutilities of work and productivities out of the conditional simulations used in the estimation process. I then undertake both nonparametric and flexible functional form regressions of work aversion 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 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[16].

Figure 9 presents the values of the maximal government receipts $K(w)$ and of the associated employment rates both for single women and for women with two children or more. As expected, government receipts and employment rates are higher for the less work averse category, here for single women without children.

Figure 10 shows the Laffer financial incentives to work for the same two categories of women (solid lines), together with the actual current incentives provided by the French taxes and social transfers to the median[17] women in the category (dotted lines). 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 competition between the various government agencies, the transfer schedule is close to maximizing government income on these two categories of women.

2. Any second best optimum yields larger incentives than the Laffer bound, which corresponds to the preferred choice of a Rawlsian planner. The proximity of the actual French system to the Laffer bound suggests that the

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

$$\omega = d + \frac{G(d)}{G'(d)},$$

and the second order condition, taken at the point ω which satisfies the first 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 ω'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.

[16]The regressors are w, its square, cube, square root and logarithm, as well as a constant term.

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

implicit social welfare criterion of French politicians, again concerning the women in the sample, is not far from Rawls.

3. This appraisal needs to be qualified: it is highly dependent on the 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.

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.

Gouriéroux, C., A. Monfort, E. Renault, and A. Trognon (1987): "Generalized Residuals," *Journal of Econometrics*, 34, 5–32.

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.

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, 205–239.

——— (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.