NBER WORKING PAPER SERIES

OPTIMAL INCOME TRANSFER PROGRAMS:
INTENSIVE VERSUS EXTENSIVE LABOR SUPPLY RESPONSES

Emmanuel Saez

Working Paper 7708
http://www.nber.org/papers/w7708

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
May 2000

I thank David Cutler, Peter Diamond, Esther Duflo, Jim Heckman, Larry Katz, Michael Kremer, John McHale, and David Spector for helpful comments and discussions. The views expressed herein are those of the author and not necessarily those of the National Bureau of Economic Research.

Optimal Income Transfer Programs: Intensive Versus Extensive
Labor Supply Responses
Emmanuel Saez
NBER Working Paper No. 7708
May 2000
JEL No. H21

ABSTRACT

This paper investigates the optimal income transfer problem at the low end of the income distribution. The government maximizes a social welfare function and faces the traditional equity-efficiency trade-off. The paper models labor supply behavioral responses along the intensive margin (hours or intensity of work on the job) and along the extensive margin (participation in the labor force). Optimal tax formulas are derived as a function of the behavioral elasticities, the shape of the income distribution and the redistribution tastes of the government. When behavioral responses are concentrated along the intensive margin, the optimal transfer program is a classical Negative Income Tax program with a substantial guaranteed income support that is taxed away at high rates. However, when behavioral responses are concentrated along the extensive margin, the optimal transfer program is an Earned Income Credit program with negative marginal tax rates at low income levels and a small guaranteed income. Numerical simulations calibrated with the actual empirical earnings distribution are presented for a range of behavioral elasticities and redistributive tastes of the government. For realistic elasticities, the optimal program provides a moderate guaranteed income, imposes low tax rates on very low annual earnings levels, and then starts phasing out benefits at substantial rates.

Emmanuel Saez
Department of Economics
Littauer Center
Harvard University
Cambridge, MA 02138
and NBER
saez@fas.harvard.edu

1 Introduction

During the twentieth century, most developed countries have adopted large and government managed income support programs. These programs have generated substantial controversy. Most recognize that they considerably improve the well-being of the disadvantaged but some have pointed out that these programs may reduce substantially the incentives to work and thus have large efficiency costs that may outweigh the redistributive gains. Redistributive public policy interventions face an equity-efficiency trade-off. Optimal transfer programs must be designed such that, at the margin, efficiency costs equate the redistributive gains. The important question is therefore to determine the optimal size and shape of the transfer program.

There is disagreement among economists and among policy makers about how income transfer programs should be structured. For a long time, economists have been inclined to think that the best transfer program is a *Negative Income Tax* (NIT) program that would provide a basic income support to all households.[1] This basic or guaranteed income would then be taxed away, but at a rate lower than 100%, as earnings or income increases until the benefit is fully lost. Many European countries use NIT type programs to redistribute toward zero or low income earners. In the U.S., there is no universal income support program. However, single headed families with dependent children are entitled to Temporary Assistance for Needy Families.[2] The TANF program works as a negative income tax program with a fairly high implicit tax rate. In many countries, the implicit tax rate at the low end of the earnings distribution is often very large because of the phasing out of transfer programs. The implicit tax rate may sometime exceed 100% when two or more transfer programs are phased-out simultaneouly.[3]

[1] For example, universal NIT program has been advocated as early as the 1940s by Milton Friedman and James Meade.

[2] The TANF program replaced in 1996 the Aid for Families with Dependent Children (AFDC) program which had been in place since 1935.

[3] Blundell and MaCurdy (1999) provide an extensive analysis of marginal tax rates faced by low income households in the U.S. and in the U.K.

A high implicit tax rate, even though it is lower than 100%, may have adverse effects on labor supply. This has lead economists and politicians to advocate programs that would make work sufficiently attractive to reduce the need for income support. These programs are known as earnings subsidies or *Earned Income Credit* (EIC) programs. An EIC does not provide any income support for individuals with no earnings but all earnings below a given threshold are partially matched by the government. In economic terms, the EIC is equivalent to negative marginal tax rates at the bottom of the income distribution. The U.S. have implemented such a program, the Earned Income Tax Credit (EITC), since 1975. In the early 1990s, because previous welfare programs such as AFDC were thought to discourage recipients to work, the EITC program was substantially increased and is now the largest cash transfer program to the poor in the U.S. There is pressure in other countries as well to move away from NIT programs toward EIC programs.[4] Obviously, relative to a NIT program, incentives to work are enhanced with an EIC program because low income workers keep a much larger share of each dollar earned than with a NIT program. However, the EIC provides no support for non-workers and thus does not reach the most needy people. Moreover, the earned income credit has also to be taxed away at some point further up the income distribution.

The main purpose of this paper is to investigate which of the NIT or the EIC programs (or which combination) is the most desirable assuming that the government values redistribution but cares about the efficiency costs of redistribution. As the efficiency costs emanate from the behavioral responses to the transfer or tax programs, it is obviously crucial to examine and model these behavioral responses as accurately as possible. The empirical literature on labor supply has emphasized two margins of labor supply responses (see Heckman (1993)). First, individuals can respond along the intensive margin. That is, they can vary their hours or intensity of work on the job. There is a very large and

[4]For example, the U.K. adopted in 1988 a Family Credit which is similar to the american EITC and that has also been expanded. Canada has implemented recently on an experimental basis a Self-Sufficiency Program that is also akin to the EITC.

controversial empirical literature on this issue (see the surveys of Pencavel (1986) and Blundell and MaCurdy (1999)). Second, individuals may respond along the extensive margin. That is, they can decide whether or not to enter the labor force. There is ample empirical evidence that the response along this margin is a real issue for low and secondary income earners. The literature consistently suggests that, at the low income end, the extensive margin of response is probably more important than the intensive margin of response.

Though the empirical literature has distinguished these two margins of behavioral responses, the implications in terms of income transfer policy have not been carefully investigated. The applied literature discusses extensively the welfare costs of actual transfer programs using behavioral response estimates from the empirical literature but does not try to derive results on optimal transfer schemes. The theoretical literature on optimal income taxation has focused almost exclusively on the intensive responses model. In that context, optimal income tax theory has been able to show that, under fairly general assumptions, optimal marginal tax rates cannot be negative. This rules out EIC programs as optimal transfer schemes.

In contrast, the main contribution of this paper is to show that the margin of the behavioral response is a key element to take into consideration when designing an optimal transfer scheme. The paper shows that a NIT program with a substantial guaranteed income level and high phasing out rates is optimal when behavioral responses are along the intensive margin. However, when behavioral responses are along the extensive margin, then the optimal transfer is similar to an EIC with negative marginal tax rates at the bottom and a smaller guaranteed income for non-workers. In the paper, optimal tax rates formulas are derived as a function of the behavioral elasticities estimated by the empirical literature. Therefore, using the estimates of both intensive and extensive behavioral elasticities from empirical studies, it is possible to assess the optimal shape and the optimal size of the transfer program. This paper provides a number of numerical simulations to investigate this point further. It is shown that four key elements affect the pattern of optimal tax rates and the size of the guaranteed income level.

First, as explained above, if the behavioral response is mostly extensive, then the guaranteed income level should be smaller and tax rates negative at the bottom so as to shift transfers away from non-workers toward low income workers. Second, the size of the behavioral elasticities at the bottom affect the optimal pattern of marginal tax rates. The higher the elasticities, the higher the efficiency costs induced by distorting prices and thus the smaller in absolute value should the size of the program be. Third, higher redistributive tastes unsurprisingly lead to a larger guaranteed income level. Last, as redistribution is financed by middle and high income earners, the higher are the behavioral responses at the middle and high income end, the higher is the cost of redistribution and thus the smaller is the optimal transfer program.

The paper is organized as follows. Section 2 reviews the main contributions of previous work on optimal transfer programs. Section 3 presents the models of intensive and extensive responses and derive optimal tax formulas in each of these two polar cases. Section 4 mixes the two polar models and derives results for that general case. Section 5 discusses the empirical literature on income maintenance programs and presents numerical simulations of optimal transfer schemes. Section 6 concludes and discusses avenues for future research.

2 Related Literature

Many economic studies have investigated the income support problem from a theoretical perspective.[5] An early literature, closely related to empirical studies, has proposed and analyzed a number of different schemes to provide income support to the disadvantaged using the classic static labor supply model. As described in the introduction, the Negative Income Tax program has been extensively discussed. Blank et al. (1999) provide a recent and detailed analysis of the effect of NIT programs on work incentives, and describe the

[5] Atkinson (1987) provides a comprehensive survey.

standard conclusion that lowering the tax rate may not necessarily reduce distortions because lowering the tax rate while keeping the same guaranteed income increases the break-even point (the income level at which benefits are fully taxed away) implying that people further up the income distribution are affected by the program.

The main other type of programs proposed are wage or earnings subsidies. The aim of these alternative programs is to make work profitable and thus reduce the need for income transfers. Kesselman (1969) and Zeckhauser (1971), among others, argue that these schemes can be more effective than NIT type programs. However, wage subsidies are difficult to implement because they depend on the wage rate, which can be manipulated more easily than total earnings. Earned income credit programs which subsidize *earnings* and which also encourage work effort are easier to implement and have been applied at a very large scale in the US through the EITC program. Wage or earning subsidies do not reach the people who do not work and have to be phased out at some point and, creating thus distortions further up the income distribution. It is therefore not obvious to assess which of the NIT or EIC program are the most appropriate to transfer income to the poor.

This literature, although it provides interesting insights into the issue of transfer programs, adopts a piecemeal approach to the income maintenance problem without specifying and solving any underlying optimization problem and does not precisely define what information the government can use when designing policy. The key issue when designing income transfer programs is precisely the impossibility to perfectly observe the individuals' earning abilities.

The theoretical literature on optimal income taxation, which grew out of the seminal contribution of Mirrlees (1971), has tried to remedy these two defects of the previous literature. However, the literature has focused almost exclusively on the intensive margin of response. Mirrlees (1971) considered such a model and characterized the optimal non-linear income tax that maximizes a given social welfare function and showed that optimal marginal tax rates at any income level cannot be negative. Seade (1982) clarified the conditions under which this result holds. Thus this important theoretical result implies that an EIC which generates negative marginal rates cannot be optimal in a model with intensive behavioral responses. Seade (1977) showed that when everybody works and when the earnings distribution is bounded away from zero then the optimal marginal rate at the very bottom is equal to zero. These conditions are unlikely to be met empirically and numerical simulations have shown that optimal rates at the bottom can be substantial when the Seade conditions do not hold (see e.g., Tuomala (1990) and Saez (2000a)). Therefore, simulations coming out of the Mirrlees (1971) model suggest that optimal transfer programs look like NIT type programs with a substantial guaranteed income taxed away at fairly high rates.

Optimal taxation models with labor force participation choice have attracted very little attention.[6] Two exceptions are Diamond (1980) and Mirrlees (1982). Diamond (1980) develops a simple optimal tax model where hours and wages are fixed but people can choose whether or not to participate in the labor force. He shows that optimal marginal tax rates may be negative for some income ranges. The study is theoretical and no attempt is made to express optimal tax formulas in terms of elasticities or to assess the importance of the participation decision margin relative to the standard intensive margin of response. As a result, that study has not been followed upon to cast light on the EIC versus NIT debate. Mirrlees (1982) is a model of optimal taxation with migration where individuals have a fixed income but can decide to migrate to other countries if taxes are too high. The decision to migrate and the decision to participate in the labor force can be modeled in almost identical terms. As a result, the optimal tax formulas obtained by Mirrlees (1982) are qualitatively close to the formulas obtained in the purely extensive case in the present paper, although their interpretation is very different because Mirrlees (1982) focuses his analysis mostly to the high end of the income distribution.

[6] Note however that retirement decision is akin to a labor force participation decision. As a result, models on the optimal shape of social security taxes and benefits, such as Diamond and Mirrlees (1978) or Diamond et al. (1980), are related to what is done in the present paper.

Last, two studies by Besley and Coate (1992, 1994) have investigated the design of income maintenance programs when the objective of the government is not to maximize social welfare but to alleviate poverty.[7] They show that in that context, work requirements (workfare) might be an effective screening device to target welfare to the less skilled individuals. Besley and Coate question the welfare maximization approach on the grounds that it requires the government to observe individuals' valuation of leisure and to make interpersonal utility comparisons. The present paper shows that it is possible to adopt a welfare maximization objective without specifying explicitly the valuation or leisure and that the redistributive tastes of the government can be summarized using straightforward marginal welfare weights at each income level.

3 Extensive versus intensive response models

In this section, I introduce the extensive model or participation model and the intensive model or effort model. I consider both discrete and continuous versions of the models. The discrete case allows simpler derivations and will be used for simulations. The continuous case allows a deeper understanding of the structure of optimal transfers. I consider income taxation at the individual level only and thus ignore completely the secondary earner labor choice decision issue that arises in the context of household income taxation.

In the discrete model, there are $I+1$ types of occupations: the unemployed earning $w_0 = 0$, and I types of jobs paying salaries w_i. The salaries w_i are increasing in i and correspond to occupations with increasing skills. The key assumption of the paper is that the government is only able to observe income levels and thus can condition taxation of income only. The net taxes paid by each class of individuals are denoted by T_i. This

[7] Kanbur, Keen and Tuomala (1994) have shown that when the objective of the government is not to maximize a classic social welfare function, but to fight poverty, negative marginal rates may be desirable for low incomes in the Mirrlees (1971) model. Their numerical simulations show, however, that this result is not important in practice and that simulated optimal fighting poverty schedules are very close to welfare maximizing simulated schedules.

tax scheme embodies both taxes and transfers. The after tax income in occupation i is denoted by $c_i = w_i - T_i$.

The total population is normalized to one and I note h_i the proportion of individuals in occupation i. Individuals choose their occupation i according to the relative after-tax rewards in each occupation. As a general formulation, the number of individuals choosing occupation i depends on after-tax rewards in all occupations: $h_i = h_i(c_0, c_1, .., c_I)$. Incentives to work are embodied in the aggregate labor supply functions h_i. High taxation and redistribution levels for example may shift labor supply away from highly productive activities toward lower productive jobs or unemployment.

The government set taxes T_i so as to maximize welfare. Taxes must finance transfers and government consumption. I assume that government consumption per capita is fixed and equal to H. The government budget constraint is,

$$\sum_{i=0}^{I} h_i T_i = H. \tag{1}$$

The welfare function can be simply characterized by marginal social welfare weights (expressed in terms of the value of public funds) that the government sets for each of the $I+1$ of occupations. These weights are denoted g_i, $i = 0, 1, .., I$ and represent the value (in terms of public funds) of giving an additional dollar to an individual in occupation i. Put another way, the government is indifferent between giving one more dollar to an individual in occupation i and g_i more dollars of public funds. As we will see, these weights are a sufficient statistic for the redistributive tastes of the government in the optimal transfer formulas that we derive. If the government values redistribution then the weights g_i are decreasing in i. In the extreme Rawlsian case where the government cares only about the worse-off individuals, all weights, except g_0, are zero.

The social weights should depend on disposable income c_i rather than w_i because the government should take into account the redistributive effects of the tax schedule when setting the marginal social weights. Therefore, it is constructive to express the social weights g_i as a function of disposable income $g_i = g(c_i)/p$ where p is the marginal value of public funds. In that case, the function $g(\cdot)$ can be taken as exogenous and it reflects the absolute redistributive tastes of the government. This formulation is useful for the numerical simulations. The function $g(\cdot)$ summarizes in a transparent way the redistributive tastes of the government and is decreasing when the government values redistribution. In that sense, welfare maximization does not require explicit interpersonal utility comparisons or to measure the value of leisure for each individual.

I assume that there are no income effects and therefore that increasing all after-tax levels c_i by a constant amount R does not change the supply levels h_i for each occupation.[8] Formally, $h_i(c_0 + R, c_1 + R, .., c_I + R) = h_i(c_0, c_1, .., c_I)$ for all i and R. With no income effects, a marginal dollar of public funds is valued as much as an additional dollar redistributed to all classes and therefore,

$$\sum_{i=0}^{I} h_i g_i = 1. \tag{2}$$

Equation (2) pins down the value of p and thus provides a normalization of the welfare weights g_i.

The model where each h_i depends on all after-tax rewards $(c_0, .., c_I)$ is of course too general to provide interesting results. Therefore, we specialize this model to two polar cases of interest: the extensive response model and the intensive response model. As is standard in the optimal income tax literature, we assume that there is perfect substitution of labor types in the production function and thus that wages w_i are fixed.[9]

3.1 Extensive Responses

- **The discrete model**

In this first model, individuals respond only through the extensive margin. Each

[8] I describe briefly below how the model can be extended to the case with income effects.

[9] Saez (2000b) has shown that the production side of the economy can be ignored when deriving optimal income tax formulas when individuals can only choose their occupation as in the model I have described.

individual is characterized by a skill level $i \in \{0, 1, .., I\}$ and may only choose either to work in occupation i corresponding to his skill or be unemployed. Therefore, the only decision is a participation decision. The decision to participate depends on the relative after-tax incomes when working c_i and when unemployed c_0. This model is obviously a crude simplification of reality but captures the extensive margin labor supply decision. As I first assume away income effects, the decision to participate depends only on the difference $c_i - c_0$. Presumably, if disposable income is higher when unemployed than when working, nobody would choose to work. Therefore, I assume that $h_i(c_i - c_0) = 0$ when $c_i \leq c_0$. As a result, it is never optimal for the government to set $c_i < c_0$ and thus I assume that $c_i \geq c_0$ for all i.

The size of the behavioral responses is captured by the elasticity of participation with respect to the difference in after-tax incomes. Formally, I define for $i = 1, .., I$,

$$\eta_i = \frac{c_i - c_0}{h_i} \frac{\partial h_i}{\partial (c_i - c_0)}. \tag{3}$$

This elasticity measures the percentage number of employed workers in occupation i who decide to leave the labor force when the difference between disposable incomes in employment and unemployment decreases by one percent. The framework underlying this elasticity is an heterogeneous population of workers (at each skill level) that are attached to the labor force in varying degrees according to their tastes for work. These tastes cannot be observed by the government and therefore taxation is based on income only.

In order to derive the optimal set of taxes T_i, I consider a small change dT_i of tax T_i on occupation i. This tax change has two effects on tax revenue and welfare.

First, there is a mechanical increase in tax revenue equal to $h_i dT_i$ because workers with skill i pay dT_i additional taxes. This increase in tax revenue, however, is valued only $(1 - g_i) h_i dT_i$ by the government because each dollar raised decreases the after tax incomes of individuals in class i and this income loss is valued g_i by the government.

Second, there is a loss in tax revenue due to the behavioral response. The small tax changes induces dh_i workers to leave the labor force. By definition of η_i in (3), we have

$dh_i = -h_i \eta_i dT_i/(c_i - c_0)$. Each worker leaving the labor force induces a loss in tax revenue equal to $T_i - T_0$, therefore the total behavioral cost is equal to $-(T_i - T_0)h_i\eta_i dT_i/(c_i - c_0)$. There is no change in welfare due to the behavioral response because workers leaving the labor force on the margin are indifferent between becoming unemployed and remaining employed.[10]

At the optimum, the sum of the mechanical and behavioral effects must be zero, implying,

$$\frac{T_i - T_0}{c_i - c_0} = \frac{1}{\eta_i}(1 - g_i). \tag{4}$$

Equations (4) for $i = 1, 2, .., I$ and (1) define the optimal set of taxes T_i for $i = 0, 1, .., I$.

With redistributive tastes, we have $g_0 > g_1 > .. > g_I$. From (2), we know that the average (using population weights) value of the g_i's is one. Therefore, there is some i^* such that $g_i \geq 1$ for $i \leq i^*$ and $g_i < 1$ for $i > i^*$. The government wants to redistribute from high skilled occupations $i > i^*$ toward low skilled occupations $i \leq i^*$.

Equation (4) implies then that $T_i - T_0 > 0$ for $i > i^*$ and that $T_i - T_0 \leq 0$ for $i \leq i^*$. When $i^* > 0$, the government provides a *higher* transfer to low skilled workers (for whom $1 \leq i \leq i^*$) than to the unemployed even though social marginal utility of consumption is highest for the unemployed. Therefore, when the government wants to redistribute toward the low income workers ($g_i > 1$ for low i), it provides them with a tax transfer $-T_i$ larger than the tax transfer to the unemployed $-T_0$. In other words, the government implements in that case a combined lumpsum guarantee income $-T_0$ and a negative marginal tax rate at the bottom (similar to the Earned Income Tax Credit) in order to increase the size of transfers as income increases. The cost of these two welfare programs are then fully financed by higher income earners.

[10]This can also be seen as a consequence of the envelope theorem: adjustments in behavior due to small price changes do not produce first order effects on utility. This property cannot be used when the objective of the government is non-welfarist. This is why welfare maximization objectives are in general easier to handle than non-welfarist objectives.

The intuition for having higher transfer levels to low skilled workers than to the unemployed is clear. Starting from a situation with equal transfers to low income workers and the unemployed, increasing the transfers to the low income workers is beneficial from a pure redistributive point of view because $g_i > 1$ for i small and positive. This also encourages some of the unemployed to join the labor force at zero fiscal cost as transfers are initially equal for the two groups. As a result, it is unambiguously welfare enhancing to provide a larger transfer to low income workers than to the unemployed.[11]

Finally, in two important cases, the EIC bubble disappears. First, when the government cares mostly about the welfare on the worse-off individuals (the extreme case being the Rawlsian objective), it might be the case that all weights (except g_0) are below one. In that case, $i^* = 0$ and $T_i \leq T_0$ for all i, implying that the negative marginal tax rate component of the welfare program disappears and the transfer program is a classic negative income tax. Second, when the government has no redistributive tastes, then there is no guaranteed income and the weights g_i are constant (below one) and set so as to raise H dollars per capita. In that case as well, tax liability is necessarily increasing with income.

- **The continuous model**

The discrete model described above can be adapted in a straightforward way to the case of a continuum of jobs. This extension is useful to investigate in more detail the optimal transfer schedule. In the continuum case, jobs are indexed by $w \in [0, +\infty)$. Job w pays a wage equal to w and job 0 corresponds to unemployment. The government implements a (non-linear) income tax schedule $T(w)$ and I note $c(w) = w - T(w)$ after-tax income. Whether an individual chooses to work or not depends on disposable income when working, $c(w) = w - T(w)$, relative to welfare benefits when unemployed which I

[11] Alternatively, using the concepts of contract theory, the intuition can be formulated as follows. In the extensive model, increasing low income salaries does not tempt higher income earners to reduce their effort to imitate low income workers but does tempt the unemployed to start working. Therefore, it is unambiguously beneficial to increase the wedge between the after-tax levels of the unemployed and of the low income workers.

note $E = c(0) = -T(0)$. I note $\hat{T}(w) = T(w) - T(0)$, the tax liability excluding the transfer E. By definition, $\hat{T}(0) = 0$.

As in the discrete model, with no income effects, the actual density of occupied jobs at income level w can be written as $h(w, \hat{c}(w))$. I also define the elasticity of participation at income level w as,

$$\eta_{(w)} = \frac{\hat{c}(w)}{h} \frac{\partial h}{\partial \hat{c}(w)}. \tag{5}$$

The optimal tax schedule is derived as above by considering a small change of tax liability $T(w)$ at income level w. The continuous version of equation (4) is,

$$\frac{\hat{T}(w)}{w - \hat{T}(w)} = \frac{1}{\eta_{(w)}} \left[1 - g(c(w))\right]. \tag{6}$$

This formula is a simple inverse elasticity tax rule. Each labor type can be considered as a good that is taxed according to the inverse of its aggregate elasticity $\eta_{(w)}$.[12] Diamond (1982) developed a fully specified particular case of the continuous model of participation just presented. He did not, however, relate his formula to participation elasticities as in equation (6).

As discussed above, $w - \hat{T}(w)$ cannot be negative because nobody would choose to work at income level w. Therefore, as $w - \hat{T}(w) \geq -\hat{T}(w)$, we have that $\hat{T}(w)/(w - \hat{T}(w))$ is larger or equal to -1. However, the right-hand side of equation (6) computed at $w = 0$ is equal to $[1 - g(c(0))]/\eta_{(0)}$ and is smaller than -1 when $g(c(0)) > 1 + \eta_{(0)}$. This happens when redistributive tastes are strong enough. In that case, there is a jump in the tax schedule at 0. The unemployed receive $c(0) = E$ and I note $c(0^+) = E + R$ the amount received by very low incomes in the limit when w tends to 0. I note $\hat{c}(0^+) = R$ the size of

[12] This formula is also closely related to the model of optimal taxation in the presence of migration developed by Mirrlees (1982). In that model, individuals can choose to work either in their country or to leave the country and work and be taxed abroad. In the present paper, the possibility of working abroad is replaced by unemployment but the structure of the model and the optimal tax formulas are almost identical.

the jump which is the extra-amount received when one starts working. For very small w, the left hand side of (6) tends to -1 and the right hand side tends to $[1 - g(R + E)]/\eta_{(0)}$. Thus, R is defined such that, $g(E + R) = 1 + \eta_{(0)}$.

Therefore, we can characterize the optimal tax schedule as follows,

1) When redistributive tastes are mild, $g(c(0)) < 1 + \eta_{(0)}$, the schedule is continuous at 0 and marginal tax rates are negative at the bottom.

2) When redistributive tastes are stronger, $g(c(0)) > 1 + \eta_{(0)}$, the schedule is discontinuous at 0. Employed individuals are paid an extra-premium R.

These two situations are illustrated on Figure 1. This graph displays two optimal tax schedules $c_H(w)$ and $c_L(w)$ corresponding respectively to High and Low redistributive tastes. The unemployed receive a positive transfer $c_H(0)$ and $c_L(0)$. In the Low redistributive tastes case, the schedule is continuous at 0 and marginal tax rates are negative at the bottom. In the High redistributive tastes case, the schedule is discontinuous at 0 ($c_L(0^+) > c_L(0)$) and low income earners receive a larger transfer than the unemployed.

In a real situation, the discontinuity is the schedule at 0 may not be desirable because of the arbitrage opportunity it can provide. Firms could agree to hire idle workers who would be paid a very small wage in order to share with the worker the government subsidy R. This arbitrage possibility issue could be alleviated by replacing the discontinuity by a large but finite earnings subsidy for very low incomes.[13] In practice, this is not an important issue as few individuals (excluding teenager dependents) report positive annual earnings below a few thousand dollars.

- **Income Effects**

In the case with income effects, the labor supply functions h_i depend not only on $c_i - c_0$ but also on c_0. Presumably, keeping $c_i - c_0$ constant, h_i is decreasing in c_0 because higher

[13]Note that any earnings subsidy provides an arbitrage opportunity for workers and firms. In practice, transaction costs and tax regulations would probably make it impossible to exploit this arbitrage opportunity.

welfare benefits make the extra-earnings derived from work less desirable. In the case with income effects, equation (2) needs to be modified. Decreasing by dT the tax liability of all individuals (including the unemployed) has a mechanical fiscal cost of dT but also induces a behavioral response through income effects. More precisely, $dh_i = dT \partial h_i / \partial c_0$ workers in occupation i shift to unemployment. The total fiscal cost due to the behavioral response is thus equal to $\sum_i (T_i - T_0) dh_i = dT \sum_i (T_i - T_0) \partial h_i / \partial c_0$. The welfare gain of the tax change is $dT \sum_i g_i h_i$. The sum of the mechanical, behavioral and welfare effects is zero at the optimum and thus we have,

$$\sum_{i=0}^{I} h_i g_i = 1 - \sum_{i=0}^{I} (T_i - T_0) \frac{\partial h_i}{\partial c_0}. \tag{7}$$

With income effects, the average welfare weights may no longer be equal to one (and is presumably larger than 1). However, the previous derivations carry over and equations (4) and (6) remain valid.

3.2 Intensive Responses

The theory of optimal income taxation has mostly focused, following Mirrlees (1971) seminal contribution, on the intensive response to taxes. The model of Mirrlees (1971) is the classical static labor supply model where individuals choose their labor supply until marginal desutility of work equals marginal utility of money derived from the extra amount of work. Key to the analysis are the elasticities of labor supply with respect to tax rates. Piketty (1997) has developed the discrete version of the Mirrlees (1971) model. He considered only the Rawlsian case but it is straightforward to adapt the model to any welfare weights. We follow here his approach.

- **The Model**

In the discrete type model, intensive responses can be modeled as follows. If the rewards of occupation i are reduced relative to the lower income occupation $i-1$, then some individuals in occupation i reduce their effort and switch to occupation $i-1$. I also assume here that there are no income effects implying that giving a uniform lumpsum to all individuals does not affect supply for each job.[14] In that case, the supply functions h_i can be written as $h_i(c_{i+1} - c_i, c_i - c_{i-1})$. When $c_{i+1} - c_i$ increases by dc and when all the other differences $c_{j+1} - c_j$ for $j \neq i$ are kept constant, there is a displacement of workers from job i to job $i+1$. The increase in h_{i+1} is exactly equal to the decrease in h_i and therefore, we have, $\partial h_{i+1}/\partial(c_{i+1} - c_i) = -\partial h_i/\partial(c_{i+1} - c_i)$. The behavioral elasticities can be defined as follows,

$$\zeta_i = \frac{c_i - c_{i-1}}{h_i} \frac{\partial h_i}{\partial(c_i - c_{i-1})}. \tag{8}$$

This elasticity measures the percentage increase in supply of job i when $c_i - c_{i-1}$ is increased by one percent. The link between this mobility elasticity and the elasticity of earnings with respect to tax rates of the usual labor supply model is investigated later on.

- **Optimal Tax Formula**

To derive optimal tax rules in this model, we consider as above a small perturbation of the optimal tax schedule. To derive a condition on the relative tax rates between jobs i and $i-1$, I consider a small increase dT in tax rates for jobs $i, i+1, .., I$: $dT_i = dT_{i+1} = .. = dT_I = dT$. This tax change decreases $c_i - c_{i-1}$ by dT but leaves unchanged all the other differences $c_j - c_{j-1}$ for $j \neq i$.

This tax change raises $[h_i + h_{i+1} + .. + h_I]dT$ additional taxes through the mechanical effect which are valued $[(1 - g_i)h_i + (1 - g_{i+1})h_{i+1} + .. + (1 - g_I)h_I]dT$ by the government. This change also induces $dh_i = -h_i \zeta_i dT/(c_i - c_{i-1})$ individuals in job i to switch to job $i-1$ reducing tax revenue by $(T_i - T_{i-1})dh_i$. At the optimum, the sum of the two effects is zero implying,

[14] Income effects can be included in the analysis as in Saez (2000a). However, income effects complicate substantially the analysis. Moreover, as income effects along the intensive margin of response have, in general, been found to be small in the empirical literature, we consider only the simpler case with no income effects.

$$\frac{T_i - T_{i-1}}{c_i - c_{i-1}} = \frac{1}{\zeta_i} \left[\frac{(1-g_i)h_i + (1-g_{i+1})h_{i+1} + .. + (1-g_I)h_I}{h_i} \right]. \tag{9}$$

Equations (1) and (9) for $i = 1, 2, .., I$ characterize the optimal tax levels T_i.

When the social weights g_i are non-increasing, equation (2) implies that $(1 - g_i)h_i + .. + (1 - g_I)h_I \geq 0$ for any $i > 0$. Therefore, formula (9) implies that tax liability T_i is *increasing* with i.[15] In the intensive model, it is therefore never optimal to impose *negative* marginal tax rates.[16] The reason for this result is the following. If there is a negative marginal rate in some range, then, by increasing slightly this rate, the government reduces work incentives in that range but this increases tax revenue raised from people in that range (precisely because the tax rate is negative in that range). Moreover, this small marginal tax rate increase allows the government to raise money from all taxpayers above that range (tax liability at any income being the sum of marginal rates up to that income). This is beneficial for redistributive reasons.[17]

The key difference between the intensive and the extensive model lies precisely in the relevant sources of incentives. In the extensive model, the behavioral response is only on the participation margin and thus what matters is only the level of taxation at each income level (relative to transfers at 0 income level). In the intensive model, the behavioral response is on the earnings level margin which implies that the relevant parameter for incentives is primarily the marginal tax rate. Note that the intensive response model generates more complicated optimal tax formulas which also depend on the shape of the income distribution. The shape of the income distribution is not directly relevant in the simple inverse elasticity rule (4).

[15]Obviously, c_i is increasing in i because nobody would choose an occupation requiring more effort and providing less after-tax income.

[16]As mentioned in Section 2, this result was noticed by Mirrlees (1971) and clarified by Seade (1982) who provides an intuition different from the one I give here.

[17]Note that when the social welfare weights are non-decreasing, negative rates might be optimal. For example, if the government cares more about low income workers (the "deserving poor") rather than the unemployed (the "undeserving poor"), then it might be optimal to provide a larger transfer to the workers than to the unemployed.

- **Relating transition elasticities to labor supply elasticities**

The discrete model of mobility that I have used does not fit directly into the Mirrlees (1971) model where individuals vary their hours of work in a continuous way. However, it can be easily shown that the two models have the same structure and that the elasticity of mobility ζ_i can be related the the usual labor supply elasticity. In order to calibrate the discrete model using the empirical elasticities from the labor supply literature, it is important to analyze the relation between the elasticities of mobility and the labor supply elasticity.

In the classical model of labor supply with no income effects, labor supply responses are measured by the elasticity of earnings with respect to one minus the marginal tax rate. This intensive elasticity of earnings is defined by,

$$\varepsilon = [(1-\tau)/w]\partial w/\partial(1-\tau). \tag{10}$$

In the discrete model, I can define an implicit marginal tax rate τ_i between occupations i and $i-1$ as, $\tau_i = (T_i - T_{i-1})/(w_i - w_{i-1})$ or equivalently $1-\tau_i = (c_i - c_{i-1})/(w_i - w_{i-1})$.

In appendix, I show that the elasticities ε_i at each income level that generate behavioral responses of the same magnitude as the elasticities ζ_i are such that,

$$\zeta_i(w_i - w_{i-1}) = \varepsilon_i w_i. \tag{11}$$

The weight h_i is the fraction of individuals at income level w_i. The discrete earnings grid should be considered as an approximation of the continuous empirical earnings density. As a result, h_i represents the number of individuals whose wages are between w_{i-1} and w_i. I can thus normalize these weights h_i by introducing $h(w_i) = h_i/(w_i - w_{i-1})$ which is the *normalized density* at income level w_i. Using this normalized density and equation (11) to replace ζ_i in the optimal tax formula (9), we obtain,

$$\frac{\tau_i}{1-\tau_i} = \frac{1}{\varepsilon_i}\left[\frac{(1-g_i)h_i + (1-g_{i+1})h_{i+1} + .. + (1-g_I)h_I}{w_i h(w_i)}\right]. \tag{12}$$

From now on, I use the labor supply elasticity ε_i when expressing optimal tax formulas.

- **Optimal rates at the bottom**

To cast light on optimal rates at the low end of the income distribution, it is useful to consider the continuous case. In the continuous limit, equation (12) becomes,[18]

$$\frac{T'(w)}{1-T'(w)} = \frac{1}{\varepsilon_{(w)} \;\; w \;\; h(w)} \int_w^\infty \left[1 - g(c(w'))\right] h(w')dw', \tag{13}$$

, Similarly, in the continuous case, formula (2) becomes,

$$\int_0^\infty g(c(w'))h(w')dw' = 1,$$

and thus, equation (13) can be rewritten as,

$$\frac{T'(w)}{1-T'(w)} = \frac{1}{\varepsilon_{(w)} \;\; w \;\; h(w)} \int_0^w \left[g(c(w')) - 1\right] h(w')dw'. \tag{14}$$

If there is no unemployment and the bottom wage w_0 is strictly positive with positive density $h(w_0)$, then the denominator in (14) tends to a positive limit while the integral term tends to zero as w tends to w_0. Therefore the optimal rate at the bottom must be zero (Seade (1977)). However, this result is no longer true when the bottom wage is equal to zero or when there is an atom non workers.

In the original Mirrlees (1971) formulation, there is a threshold skill level under which individuals do not work. This implies that the intensive elasticity at the bottom is infinite. Therefore, there is an element of labor force participation in the intensive model of Mirrlees (1971). But the labor force participation choice is only between unemployment and an infinitesimal amount of work. This feature is not empirically realistic, because of fixed costs of work; the model developed in Section 3.1 is closer to the actual economic situation. Within the framework of this paper, it is conceptually useful to distinguish clearly the extensive response model from the intensive response model. Therefore, we consider here

[18] This equation was developed by Diamond (1998) and Saez (2000a).

the pure intensive model with no participation element whatsoever. Consequently, we assume that everybody works and that the intensive elasticity is finite at the bottom.[19]

Assuming that the bottom wage is zero and that $h(w)$ tends to a finite limit $h(0^+)$ when w tends to zero, using equation (14), we obtain in the limit,

$$\frac{T'(0)}{1-T'(0)} = \frac{g(c(0))-1}{\varepsilon_{(0)}}. \tag{15}$$

Therefore, the higher the redistributive tastes of the government at the bottom, the higher the optimal rate at the bottom.[20] The intuition for this result is that the best way to redistribute to zero or low income earners, is by making the lumpsum transfer $-T(0)$ as large as possible, this can be achieved only by imposing large rates at the bottom. Note also that, as usual, the optimal rate is decreasing the the elasticity level $\varepsilon_{(0)}$ because high elasticity levels imply that tax rate distortions have high efficiency costs. As an example, consider the case where $\varepsilon_{(0)} = 0.5$ and where $g(c(0)) = 3$. That is, the government values 3 times as much a one dollar transfer to a worker with income close to zero than a three dollar transfer uniformly spread over all individuals. In that case, equation (15) implies that the optimal rate at the bottom is 80%. Saez (2000a) presents a number of simulations in the pure intensive model. The optimal rates obtained at the bottom are fairly high, in general in excess of 70%.

The general conclusion of the pure intensive model is that redistribution should take place through a guaranteed income level that should be taxed away as income increase. The transfer program takes the form of a traditional Negative Income Tax and no Earned Income Credit component should be included.

[19] This particular case has not been studied in detail and thus the formula I give below has not, to the best of my knowledge, been obtained before.

[20] In the extreme Rawlsian case where $g(c(0))$ is infinite, the optimal rate at the bottom is one.

4 A General Model with Extensive and Intensive Responses

The previous Section has illustrated the contrast between the intensive and extensive model for designing an optimal transfer scheme. However, the real world is obviously a mix of the two models. The goal of this section is thus to develop a model which incorporates both the extensive and the intensive margin of response in order to assess how these two effects interact and to perform numerical simulations.

In the pure extensive response model, the number of individuals in occupation i depends on $c_i - c_0$. In the pure intensive model, the number of individuals in occupation i depends on $c_i - c_{i-1}$ and on $c_{i+1} - c_i$. In the general model with both extensive and intensive responses, the supply in each job i is given by $h_i(c_i - c_0, c_{i+1} - c_i, c_i - c_{i-1})$.

The optimal tax formulas in the mixed model are derived in appendix using the same methodology as in Section 3. The optimal tax formula expressed in terms of the labor supply elasticities ε_i and the participation elasticities η_i is given by,

$$\frac{\tau_i}{1-\tau_i} = \frac{1}{\varepsilon_i \;\, w_i \;\, h(w_i)} \sum_{j \geq i} h_j \left[1 - g_j - \eta_j \frac{T_j - T_0}{c_j - c_0} \right]. \tag{16}$$

Comparing equations (12) and (16), we see that the mixed model is identical to the intensive model with weights g_j replaced by $\hat{g}_j = g_j + \eta_j (T_j - T_0)/(c_j - c_0)$. Therefore, adding the participation margin amounts to attributing a higher welfare weights $\hat{g}_j$ to income groups that are prone to leave the labor force and that receive a lower transfer than the unemployed ($T_j > T_0$). It is also possible to derive the optimal tax formula in the continuous case (see appendix).

Each of the two polar cases analyzed in Section 3 can be obtained from equations (16) or (22) by letting either one of two elasticities (η or ε) tend to zero. As a result, if the participation elasticity is large relative to the earnings elasticity, the optimal schedule will have an EIC component with larger transfers for low income workers than for the unemployed. More precisely, because of the extensive margin of response, the pseudo weights $\hat{g}_i$ need not be decreasing even if the real weights g_i are decreasing. As a result, optimal tax rates are not necessarily non-negative as in the pure intensive model. On the other hand, if the elasticity η is small relative to ε, then the optimal schedule will have non-negative rates everywhere as in the standard intensive model.

5 Empirical Calibration

5.1 Empirical Literature

As mentioned in Introduction, the empirical literature on labor supply and behavioral responses to taxes and transfers is large. Hausman (1985), Pencavel (1986), and Blundell and MaCurdy (1999) provide extensive reviews of the literature. Most studies find that the intensive labor supply elasticities of males are small. However, elasticities of labor force participation have been found to be much larger for some classes of the population such as the elderly, single mothers or secondary earners. It should be noticed, though, that studies often do not separate participation and intensive elasticities. For example, many structural studies estimate substitution and income effects and do not provide specific estimates on the participation margin of response.

For our simulations, we need to pay special attention to elasticities of earnings and participation at the bottom of the income distribution. Four pieces of evidence are of particular interest.

First, in the late 1960s, a series of Negative Income Tax experiments were implemented in the U.S. These experiments provide in principle an ideal set-up to estimate both participation and intensive elasticities of labor supply. Robbins (1985) surveys the empirical results based on NIT experiments. Both intensive and extensive elasticities for males are small (around 0.2). The behavioral response for wives, single female heads and the young is higher and concentrated along the participation margin. Participation elasticities are often in excess of 0.5 and sometimes close to 1.[21]

[21] Ashenfelter (1978), for example, using data from the North Carolina-Iowa Rural Income Maintenance

Second, a number of studies have used actual transfer programs in the U.S. (mostly AFDC) to estimate behavioral responses (see the surveys by Danziger et al. (1981) and Moffitt (1992)). However, most studies regress labor supply hours or participation on parameters of the welfare programs such as the guaranteed income level, the phasing-out rate, or the break-even level. As a result, it is difficult to compare those estimates to intensive and extensive elasticities used in the present paper. Moreover, there is substantial disagreement about the size of the effects. Nevertheless, recent studies exploiting the recent increases in the EITC (see e.g. Eissa and Liebman (1996), Meyer and Rosenbaum (1998)), have shown that the effect on participation of single female heads is substantial.[22]

Last, following the recent discontent with the classical NIT programs, some U.S. states and Canada have carried out experimental welfare programs with lower implicit tax rates or with minimum weekly hours of work requirements (see Blank et al. (1999)). In Canada, in particular, a large experiment called the Self-Sufficiency Project (SSP) has been implemented. SSP increased substantially the wedge between disposable income when working and when non-working. Card and Robbins (1996) estimate large effects on employment which suggests participation elasticities between 0.5 and 1 for current welfare recipients.

To summarize, the literature suggests that participation elasticities at the low end of the income distribution maybe large (perhaps above 0.5). Elasticities of earnings with respect to the tax rate are substantially smaller (perhaps around 0.25). The elasticity of participation at the middle and high end of the income distribution is very likely to be small. There is little consensus about the magnitude of intensive elasticities of earnings for middle income earners, though this elasticity is likely to be of modest size for middle income earners and higher for high income earners. Gruber and Saez (2000) summarize this literature and display empirical estimates between 0.25 and 0.5 for middle and high income earners.

Experiment, reports elasticities for wives around 0.9 while elasticities for husbands are only around 0.2.

[22]For example, Eissa and Liebman (1996) report participation elasticities for single mothers with low education around 0.6.

5.2 Numerical Simulations

- **Calibration**

Numerical simulations are based on the discrete model mixing the extensive and intensive margin of behavioral response. The discrete model is preferred because it is far easier to implement numerically while very little is lost by replacing a continuum of incomes by a sufficiently high number of income groups. In the simulations, I use a discrete grid of 17 income levels (see Table A1). I describe in appendix the technical details of the simulations. A number of parameters are crucial for tax schedule simulations.

First, the elasticity parameters summarizing the behavioral responses are prominent. As there is no strong consensus on the size of these parameters, I present simulations using a range of plausible parameter values. The values for the participation elasticity η has been taken as constant and equal to 0, 0.5 or 1 for incomes below \$20,000 and equal to 0 for incomes above \$20,000. The intensive elasticity ε_L for incomes below \$20,000 is taken as constant and equal to 0, 0.25 or 0.5. The middle and high income (above \$20,000) elasticities ε_H is taken as constant and equal to 0.25 or 0.5. All simulations have been carried out assuming no income effects.[23]

Second, the social welfare weights which summarize the redistributive tastes of the government may also affect the optimal level and patterns of taxation. I summarize the redistributive tastes of the government using a simple parametric form for the curve of marginal weights $g(c) = 1/(p \cdot c^{\nu})$ where p denotes the marginal value of public funds and ν is a scalar parameter. The higher is ν, the higher are the redistributive tastes of the government. $\nu = +\infty$ corresponds to the Rawlsian criterion while $\nu = 0$ corresponds no redistributive tastes. Most of the simulations are presented with $\nu = 1$ which represents fairly strong redistributive tastes. With $\nu = 1$, the government values N times less marginal consumption when disposable income is multiplied by N.

[23]Simulations including income effects in the pure extensive model of Section 3.1 have been performed. Income effects have little effect of the size and shape of the optimal transfer program but can have a substantial effect on the percentage of unemployed workers.

Third, the income distribution is calibrated using the empirical yearly earnings distribution from the March 1997 Current Population Survey (CPS). I limit the sample to individuals aged 18 to 60 and I exclude students. The details are explained in appendix. The rate of non-labor force participation (zero yearly earnings reported) for this group is slightly above 15%.

Last, the exogenous revenue requirement of the government H obviously affects the general levels of taxation and transfers. I assume that the government wants to collect the same amount that is actually collected with the income tax (state and federal) net of redistribution done with the earned income tax credit. This amount is around $8,000 per household, implying an average annual tax per adult around $5,000. Therefore, in the simulations, H is taken as equal to $5,000. Therefore, the tax schedules presented are roughly comparable to the actual welfare and income tax schedule.

- **Results**

The results of numerical simulations are presented in the four panels of Figure 2 and in Tables 1 and 2. Figure 2 displays optimal disposable income schedules $c(w) = w - T(w)$ as a function of earnings w. As the emphasis is on the low income end, the graphs are plotted for the income range $0 to $20,000. The 45^o degree line gives the benchmark schedule with no tax or subsidy. Each figure displays the optimal schedules for three values of the participation elasticity η (0, 0.5 and 1) and fixed values of the intensive elasticities for high and low incomes (ε_L and ε_L) and the redistributive tastes (parameter ν).Income effects are assumed away in all simulations.

In the top-left panel, $\varepsilon_L = 0.25$, $\varepsilon_H = 0.25$ and $\nu = 1$ and there are no income effects. This panel shows that increasing η affects substantially the shape of the optimal transfer program. With $\eta = 0$, the program is a traditional NIT with a substantial guaranteed income ($9,900) and high phasing out rates (around 70%). However, when η increases, an EIC bubble appears at the low end. The guaranteed income is reduced ($4,500 for $\eta = 1$) and earnings are slightly subsidized over the income range $0 to $6,000. From $6,000 to

$15,000, the transfer is taxed away at a high rate (in excess of 50%).

The top-right panel of Figure 2 displays the same graphs with a higher intensive elasticity for low incomes $\varepsilon_L = 0.5$. Increasing ε_L, reduces the size of the EIC bubble. It also decreases slightly the guaranteed income level and decreases the phasing out rate.

The bottom panels of Figure 2 consider variations in the redistributive parameter ν. In the bottom-left panel, $\nu = 4$ which represents extremely strong redistributive tastes close to the Rawlsian case.[24] Relative to the top-left panel benchmark, the size of the guaranteed income is substantially higher (around $12,000), the phasing out rates is very high (around 80%), and the EIC bubble has completely disappeared: the elasticity of participation η has little effect on the optimal schedule. In the bottom-right panel, $\nu = 0.25$ which represents a very low taste for redistribution.[25] Relative to the top-left panel, the guaranteed income is very small (less than $2,000 for $\eta = 0.5, 1$). Both the EIC bubble and the phasing out rates are small.

Tables 1 and 2 summarize the optimal schedules for a wide range of parameters. Optimal schedules are summarized by 5 numbers. First, the guaranteed income level $-T(0)$ that is provided to the unemployed. Second, the average marginal tax rate from $0 to $6,000 ($[T(6000) - T(0)]/6000$) measures the tax distortion at the lowest end of the earnings distribution. Third, the average marginal tax rate from $6,000 to $15,000 ($[T(15000) - T(6000)]/[15000 - 6000]$) measures the phasing out rate of the transfer program. Fourth, the break-even point is the income level at which transfers are equal to zero ($T(w) = 0$) and the consumption schedule $c(w)$ crosses the 45^o line. Fifth, the average marginal tax rate from $30,000 to $100,000 measures the tax burden for middle and high income earners. Finally, the level of unemployment induced by the optimal transfer program is reported.

Table 1 focuses on the case where the redistributive taste parameter ν is equal to one.

[24] With $\nu = 4$, the government values 16 times less marginal consumption when disposable income doubles.

[25] With $\nu = 0.25$, the government values only 2 times less marginal consumption when disposable income is multiplied by 16.

In the left side of the table (columns (1) to (6)), $\varepsilon_H = 0.25$ while in the right side, (columns (7) to (12)), $\varepsilon_H = 0.5$. Table 1 confirms the graphical results. Unsurprisingly, a higher elasticity ε_H reduces the optimal rates for high incomes and thus the size of the optimal transfer program. The simulations show that, when η is high, the level of transfers affects significantly the unemployment rate and hence the cost of the transfer program. With $\eta = 1$, the optimal transfer program reduces the non-labor force participation rate to less than 3% which is only a fifth of the baseline level of non labor force participation. It is clearly unrealistic to assume that changing the parameters of the tax schedule can have such a large effect on labor force participation. Therefore, if the participation elasticity η evaluated around the current tax schedule is really equal to one, it seems likely that this elasticity is going to decrease as more and more people are induced to join the labor force and as only the individuals who really do not want to work are left in the pool of non-participants. In other words, though formula (16) is perfectly valid even if elasticities vary with the program parameters, it is no longer appropriate to use current elasticities estimates and apply them to a tax situation very different from the present one.[26]

Table 2 investigates the effect of the redistributive taste ν. The cases $\nu = 0.25$, $\nu = 1$ and $\nu = 4$ are considered in Panels A, B and C. Unsurprisingly, higher redistributive tastes lead to a larger guaranteed income level and higher phasing out rates and higher rates for middle and high incomes.

From the empirical literature, we can consider the case $\eta = 0.5$, $\varepsilon_L = 0.25$ and $\varepsilon_H = 0.25$ as a plausible benchmark. In that case with $\nu = 1$, the optimal transfer program should consist in a guaranteed income level with a modest tax rate for the first few thousands dollars of earned income (tax rate around 10% for the first \$4,000 earned). The transfer income should then be taxed at fairly high rates further up the income distribution (tax rate around 60% from \$4,000 to \$15,000). Tax rates should then be lower (around 50%) for middle and high income earners. The size of the guaranteed income level

[26]This illustrates the well known difficulty to perform out of sample predictions using empirical estimates that are in principle valid only in the sample.

is relatively large at \$7,300. Therefore, the simulations suggest that combining a sizeable Negative Income Tax program with a tax exemption for the first five thousands dollars of annual earnings might be a desirable way to redistribute toward the disadvantaged. The guaranteed income provides income support for the really needy, and the earnings exemption does not discourage too severely work participation of people with low earnings potential.

- **Comparing Simulations to the Current Transfer System**

The model used in the paper is of course a crude approximation to the actual economic situation and therefore simulations should be regarded as illustrative only. It is nonetheless interesting to speculate what type of elasticity parameters could justify the current structure of the U.S. transfer program policy. The current U.S. system applies very different tax schedules depending on the family status of the households. Two parents low income families are not in general eligible for welfare programs and thus can only collect EITC benefits implying that they face negative marginal rates of -32% (including Social Security Employee Payroll Taxes) over the first \$10,000 dollars of household earnings and tax rates around 30% on the next \$10,000 of earnings. Participation elasticities for two parent families should be very large to justify such a large program. Note that there is no universal program providing a minimum guaranteed income to these type of families with no earnings at all.

On the other hand, single parent families are also entitled to TANF and Food Stamps which are NIT type programs. This programs provide a basic support (around \$10,000 but which large variations across states). The sum of Individual income taxation, EITC, TANF and food stamps generates tax rates around 20% for the first \$9,000 of earnings but much higher rates, around 70%, for the next \$9,000 of earnings. As discussed above, most estimates suggest that participation elasticities for these group are large (in excess of 0.5). It is striking to see how close this schedule is to our benchmark simulation result ($\eta = 0.5$, $\epsilon_H = \epsilon_L = 0.25$, and $\nu = 1$).

Finally, families or individuals with no children are entitled to very little benefits. This is a well-known gap in the U.S. welfare structure that is not justified in the type of model we have considered. Therefore, the current welfare U.S. comes fairly close to our simulation results only for single parents families. This group, however, represents a large fraction of the population in need of income support.

6 Conclusion

This paper has shown that the nature of labor supply responses to taxes and transfers is critical to design optimal income transfer programs. If the behavioral response is mainly along the intensive margin then the optimal program is a classical Negative Income Tax program with a large guaranteed income level which is taxed away at high rates. However, if the behavioral response is concentrated along the extensive or labor force participation margin then the optimal program is an Earned Income Credit with a smaller guaranteed income level and transfers that *increase* with earnings at low income levels. Formulas for optimal tax rates have been derived in terms of the behavioral elasticities and the redistributive tastes of the government.

The main lesson from the numerical simulations is that the optimal program is fairly sensitive to the size of the participation elasticity. When the participation elasticity is zero, the optimal program is a large Negative Income Program with a guaranteed income in excess of $10,000 and a high phasing-out rate (around 70%). However, if the participation elasticity is substantial, then the guaranteed income level should be lower but the first $5,000 to $7,000 should be exempted from taxation (or even slightly subsidized). The guaranteed income should then be taxed at a fairly high rate for incomes between $6,000 and $15,000. It is therefore critical to distinguish carefully participation and intensive elasticities in empirical studies. If the participation elasticity is large, then very strong redistributive tastes are needed to obtain an optimal guaranteed income level above the poverty level. The combined EITC and welfare U.S. system for single mothers is close to our optimal simulated schedules if, as evidenced by empirical studies, participation elasticities are substantial.

The present model could be extended in three directions. First, the paper considered a model of individual labor supply decisions. An important feature that is missing is the secondary earner labor supply decision. There is ample empirical evidence that the labor participation decision of wives is very elastic. This suggests that a tax on total household earnings as in the U.S. might be inefficient. Eissa and Hoynes (1998) argue that the EITC program which is based on total earnings in the household discourages female labor participation. At the same time, a tax on individuals, as in the U.K., is more efficient but also less equitable because total household income is a better indicator of well-being than individual income. The secondary earner problem raises an interesting and difficult optimal income tax problem which could be tackled using the methods developed in this paper. Note also that participation elasticities is likely to be correlated with fixed costs of work. As a result, single headed families with young kids for example are more likely to be very elastic on the participation margin and should be encouraged to work through EIC type programs.

Second, there is evidence in the labor literature that long term unemployment experiences may have an adverse effect on human capital and thus on subsequent wages. This problem is especially acute in Europe. This extra cost of unemployment has not been taken into account in the present paper. Plausibly, if this extra cost is high, then the optimal policy should be tilted even more toward EIC type programs and away from NIT programs. An important element to consider when designing the optimal policy in this case is whether individuals fully internalize the extra cost of unemployment.

Third, using empirical elasticities, it would be interesting to infer the social weights g_i that make the actual US tax and transfer system optimal. Even if the government does not explicitly maximize welfare, it may be interesting to know what are the implicit weights that the government is using. For example, if some of the weights appear to be negative then the tax schedule is not second-best Pareto efficient.[27]

Last, this study has been carried out in a timeless economy and has ignored the important question of the time period on which tax liability is computed. This implicit time period obviously affects the relative scope of extensive versus intensive margin. If the time period is a lifetime, then, as almost every person does some work over his lifetime, the extensive margin becomes irrelevant and the behavioral responses are necessarily intensive. On the other hand, if the time period is very short, such as a day or an hour, the extensive margin becomes prominent. Therefore, introducing time raises the important but difficult question on the optimal period that should be taken into account to compute tax liability. Relatively little work has been done on this subject.[28] It might be the case that the time dimension for assessing optimal tax and transfers programs is important and that applying EIC programs on a monthly or quarterly basis could be more effective than an annual basis. This largely under-explored issue is left for future research.

[27]This analysis has been used frequently in the commodity taxation literature where it is known as the inverse optimum problem (Ahmad and Stern (1984)) but has never been applied to the transfer program problem.

[28]Vickrey is the economist who has studied the issue the most carefully. He advocated a system of life-time taxation (see Vickrey (1947)).

Appendix

- **Relating ζ_i and ε_i**

Consider as seen in Section 3.2, a small change dT in all tax levels T_j for $j \geq i$. By definition of the implicit marginal tax rate τ_i, this tax change is equivalent to a change in the marginal rate τ_i equal to $d\tau_i/(1-\tau_i) = dT/(c_i - c_{i-1})$.

Using the mobility elasticity ζ_i, this small tax change induces a loss in tax revenue equal to,

$$-(T_i - T_{i-1})h_i\zeta_i \frac{dT}{c_i - c_{i-1}} = -\tau_i h_i (w_i - w_{i-1})\zeta_i \frac{d\tau_i}{1-\tau_i}. \tag{17}$$

In the classic labor supply supply model, by definition of the earnings elasticity ε_i, this tax change reduces earnings of the individuals with income w_i by $dw = -\varepsilon_i w_i d\tau_i/(1-\tau_i)$. As there are h_i individuals with income w_i, the total effect on tax revenue is equal to $\tau_i h_i dw$ which can be written as,

$$-\tau_i w_i h_i \varepsilon_i \frac{d\tau_i}{1-\tau_i}. \tag{18}$$

Therefore, comparing (17) and (18), we see that the two models produce the same behavioral response when,

$$\zeta_i(w_i - w_{i-1}) = \varepsilon_i w_i. \tag{19}$$

- **Optimal Tax Formula in the General Model**

The optimal tax formula can be derived as in Section 3.2 by considering a small change dT in the tax rates for jobs $i, i+1, .., I$: $dT_i = dT_{i+1} = .. = dT_I = dT$. This tax change raises $[h_i + h_{i+1} + .. + h_I]dT$ additional taxes through the mechanical effect which are valued $[(1-g_i)h_i + (1-g_{i+1})h_{i+1} + .. + (1-g_I)h_I]dT$ by the government.

As in Section 3.2, this tax change decreases $c_i - c_{i-1}$ by dT and leaves unchanged all the other differences $c_j - c_{j-1}$ which induces $dh_i = -h_i\zeta_i dT/(c_i - c_{i-1})$ individuals in job i to switch to job $i-1$ reducing tax revenue by $(T_i - T_{i-1})dh_i = -(T_i - T_{i-1})h_i \zeta_i dT/(c_i - c_{i-1})$.

This tax change also changes all the differences $c_j - c_0$ for $j \geq i$ and thus induce a number $-h_j \eta_j dT/(c_j - c_0)$ of individuals in each occupation $j \geq i$ to become unemployed. Therefore the total behavioral cost due to movements in and out unemployment is equal to $-dT \sum_{j \geq i}(T_j - T_0)h_j \eta_j/(c_j - c_0)$.

At the optimum, the sum of these effects is zero, implying,

$$\frac{T_i - T_{i-1}}{c_i - c_{i-1}} = \frac{1}{\zeta_i \ h_i} \sum_{j \geq i} h_j \left[1 - g_j - \eta_j \frac{T_j - T_0}{c_j - c_0} \right]. \tag{20}$$

Using (11), we can express this formula using the usual labor supply elasticities ε_i and the normalized density $h(w_i)$ introduced in Section 3.2,

$$\frac{\tau_i}{1 - \tau_i} = \frac{1}{\varepsilon_i \ w_i \ h(w_i)} \sum_{j \geq i} h_j \left[1 - g_j - \eta_j \frac{T_j - T_0}{c_j - c_0} \right]. \tag{21}$$

- **Optimal Tax Formula in the Continuous Model**

The optimal tax formula takes the following form,

$$\frac{T'(w)}{1 - T'(w)} = \frac{1}{\varepsilon_{(w)} \ w \ h(w)} \int_w^{\infty} \left[1 - g(c(w')) - \eta_{(w')} \frac{\hat{T}(w')}{w' - \hat{T}(w')} \right] h(w')dw'. \tag{22}$$

- **Numerical Simulations**

The numerical simulations are performed using the empirical earnings distribution. The data used to calibrate the earnings distribution is annual individual earnings data from the March 1997 Current Population Survey. The simulations are performed using the discrete model of Section 4. Therefore, the empirical earnings distribution is approximated using a discrete grid. The vector of discrete values for earnings levels is reported in Table A1. The vector of density weights h_i^0 estimated using data on the empirical distribution is also reported in Table A1. The non labor force participation rate is equal to 15%.

The system consists in $I + 2$ simultaneously equations (2), (1), (16) for $i = 1, .., I$. The welfare weights are $g_i = 1/(p \cdot c_i^{\nu})$ where p is the marginal value of public funds and

ν is the redistributive tastes parameter. There are $I+2$ unknowns, the tax levels T_i for $i = 0, 1, .., I$ and the marginal value of public funds p. The system has $I+2$ equations and $I+2$ unknowns and thus yields in practice a unique solution.

The main complication of simulations comes from the endogeneity of the density weights h_i. The density weights h_i are endogenous because the distribution of earnings and the unemployment level are affected by taxes and transfers. Formally, Section 4 has shown that the functional form of the density weights is $h_i(c_i - c_0, c_{i+1} - c_i, c_i - c_{i+1})$. In principle, the weights h_i should satisfy two conditions. First, the functional form of the weights h_i should be chosen so as to be compatible with the structure of behavioral elasticities η_i and ζ_i defined in equations (3), and (8)). Second, the weights h_i should coincide with the empirical weights h_i^0 when the tax schedule $(T_i, i = 0, 1, .., I)$ is equal to the actual schedule $(T_i^0, i = 0, 1, .., I)$.

However, it is impossible to find functions $h_i(c_i - c_0, c_{i+1} - c_i, c_i - c_{i+1})$ that satisfy equations (3), and (8) for constant elasticities η_i, ζ_i and δ_i for all possible values of c_0 and $c_i, i = 0, 1, .., I$. Therefore, in the simulations, I ignore the effect of the intensive behavioral response on h_i. The density weights are taken as,

$$h_i = h_i^0 \cdot \left(\frac{c_i - c_0}{c_i^0 - c_0^0} \right)^{\eta_i},$$

where $c_i^0, I = 0, 1, .., I$ is the actual after-tax schedule. The schedule c_i^0 used in simulations is a very simplified approximation of the real schedule. The real schedule is approximated with a linear tax schedule with constant tax rate of 40% and a guaranteed income $c_0^0 =$ \$6,000. Sensitivity analysis shows that the optimal schedules are not significantly affected when other assumptions for the actual schedule c_i^0 are made.

References

[1] Ahmad Ehtisham and Nicholas Stern, "The Theory of Reform and Indian Indirect Taxes", Journal of Public Economics, 25(3), 1984, pp. 259-298.

[2] Ashenfelter, Orley, "The Labor Supply Response of Wage Earners", in J. Palmer and J. Pechman, eds., *Welfare in Rural Areas*, Washington, D.C.: The Brookings Institution, 1978.

[3] Atkinson, Anthony B. "Income Maintenance and Social Insurance." in A.J. Auerbach and M. Feldstein (eds.), *Handbook of Public Economics*, Volume II, Amsterdam: North-Holland, 1987.

[4] Besley, Timothy and Stephen Coate. "Workfare versus Welfare: Incentive Arguments fro Work Requirements in Poverty Alleviation Programs." *American Economic Review*, 1992, 82(1), pp. 249-61.

[5] Besley, Timothy and Stephen Coate. "The Design of Income Maintenance Programmes." *Review of Economic studies*, 1994, 62, pp. 187-221.

[6] Blank, Rebecca M., Card David and Philip K. Robins. "Financial Incentives for Increasing Work and Income Among Low-Income Families.", NBER Working Paper No. 6998, 1999.

[7] Blundell, Richard and Thomas MaCurdy, "Labor Supply: A Review of Alternative Approaches", in O. Ashenfelter and D. Card, eds., *Handbook of Labor Economics*, Volume 3A, 1999.

[8] Card, David and Philip K. Robins. "Do Financial Incentives Encourage Welfare Recipients to Work? Evidence from a Randomized Evaluation of the Self-Sufficiency Project.", NBER Working Paper No. 5701, 1996.

[9] Danziger, Sheldon, Robert Haveman and Robert Plotnick. "How Income Transfer Programs Affect Work, Savings, and the Income Distribution: A Critical Review.", *Journal of Economic Literature*, 19, 1981, pp. 975-1028.

[10] Diamond, Peter. "Income Taxation with Fixed Hours of Work." *Journal of Public Economics*, 1980,13, pp. 101-110.

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

[12] Diamond, Peter, L. Helms and James Mirrlees. "Optimal Taxation in a Stochastic Economy: A Cobb-Douglas Example." *Journal of Public Economics*, 1980, 14, pp. 1-29.

[13] Diamond, Peter, and James Mirrlees. "A Model of Optimal Social Insurance with Variable Retirement." *Journal of Public Economics*, 1978, 10, pp. 295-336.

[14] Eissa, Nada and Hilary Hoynes. "The Earned Income Tax Credit and the Labor Supply of Married Couples.", NBER Working Paper No. 6856, 1998.

[15] Eissa, Nada and Jeffrey Liebman. "Labor Supply Response to the Earned Income Tax Credit.", *The Quaterly Journal of Economics*, 111, 1996, pp. 605-37.

[16] Gruber, Jonathan and Emmanuel Saez, "The Elasticity of Taxable Income: Evidence and Implications.", NBER Working Paper No. 7512, 2000.

[17] Hausman, Jerry, "Taxes and Labor Supply." in A.J. Auerbach and M. Feldstein (eds.), *Handbook of Public Economics*, Amsterdam: North-Holland, 1985.

[18] Heckman, James. "What has been learned about Labor Supply in the Past Twenty Years?", *American Economic Review*, 1993, 83(2), pp. 116-121.

[19] Kanbur, Ravi, Keen, Michael and Tuomala, Matti. “Optimal Non-linear Income Taxation for the Alleviation of Income-Poverty.” *European Economic Review*, 38(8), 1994, pp. 1613-32.

[20] Kesselman, Jonathan. “Labor-Supply Effects of Income, Income-Work, and Wage Subsidies.” *Journal of Human Resources*, 4(3), 1969, pp. 275-92.

[21] Meyer, Bruce and Daniel Rosenbaum. “Welfare, the Earned Income Tax Credit, and the Labor Supply of Single Mothers.”, NBER Working Paper No. 7363, 1999.

[22] Mirrlees, James A. “An Exploration in the Theory of Optimal Income Taxation.”, *Review of Economic studies*, 1971, 38, pp. 175-208.

[23] Mirrlees, James A. “Migration and Optimal Income Taxes.”, *Journal of Public Economics*, 1982, 18(3), pp. 319-41.

[24] Moffitt, Robert. “Incentive Effects of the U.S. Welfare System: A Review.”, *Journal of Economic Literature*, 30, 1992, pp. 1-61.

[25] Pencavel, John. “Labor Supply of Men.” in O. Ashenfelter and R. Layard (eds.), *Handbook of Labor Economics*, 1986, Amsterdam: North-Holland, pp. 3-102.

[26] Piketty, Thomas. “La Redistribution Fiscale face au Chômage.” *Revue Française d’Economie*, 1997, 12(1), pp. 157-201.

[27] Robins, Philip K. “A Comparison of the Labor Supply Findings from the Four Negative Income Tax Experiments.” *Journal of Human Resources*, 20, 1985, pp. 567-82.

[28] Saez, Emmanuel. “Using Elasticities to Derive Optimal Income Tax Rates.”, NBER Working Paper No. 7628, 2000a.

[29] Saez, Emmanuel. “Optimal Income Taxation, Production Efficiency and Tariffs: Short-Run Versus Long-Run.”, Harvard mimeograph, 2000b.

[30] Seade, Jesus K. "On the Shape of Optimal Tax Schedules." *Journal of Public Economics*, 1977, 7(2), pp. 203-236.

[31] Seade, Jesus K. "On the Sign of the Optimum Marginal Income Tax." *Review of Economic Studies*, 49(4), 1982, pp. 637-643.

[32] Tuomala, Matti. *Optimal Income Tax and Redistribution.* Oxford: Clarendon Press, 1990.

[33] Vickrey, William. *Agenda for Progressive Taxation*, The Ronald Press Company: New York, 1947.

[34] Zeckhauser, Richard J. "Optimal Mechanisms for Income Transfer.", *American Economic Review*, 61(3), 1971, pp. 324-34.

Figure 1: Optimal Transfer Schedules, Extensive Margin

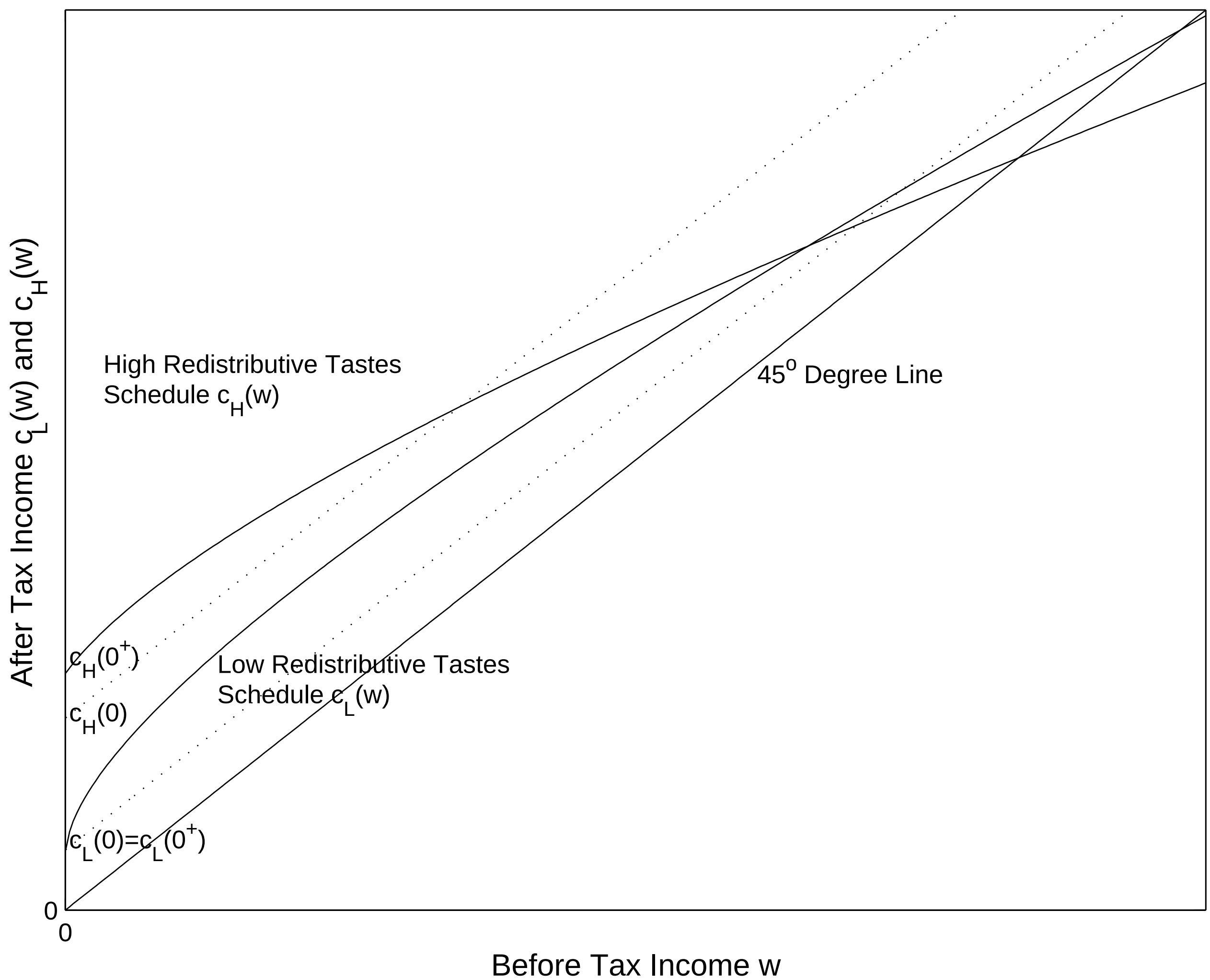

Figure 2. Optimal Tax Schedules

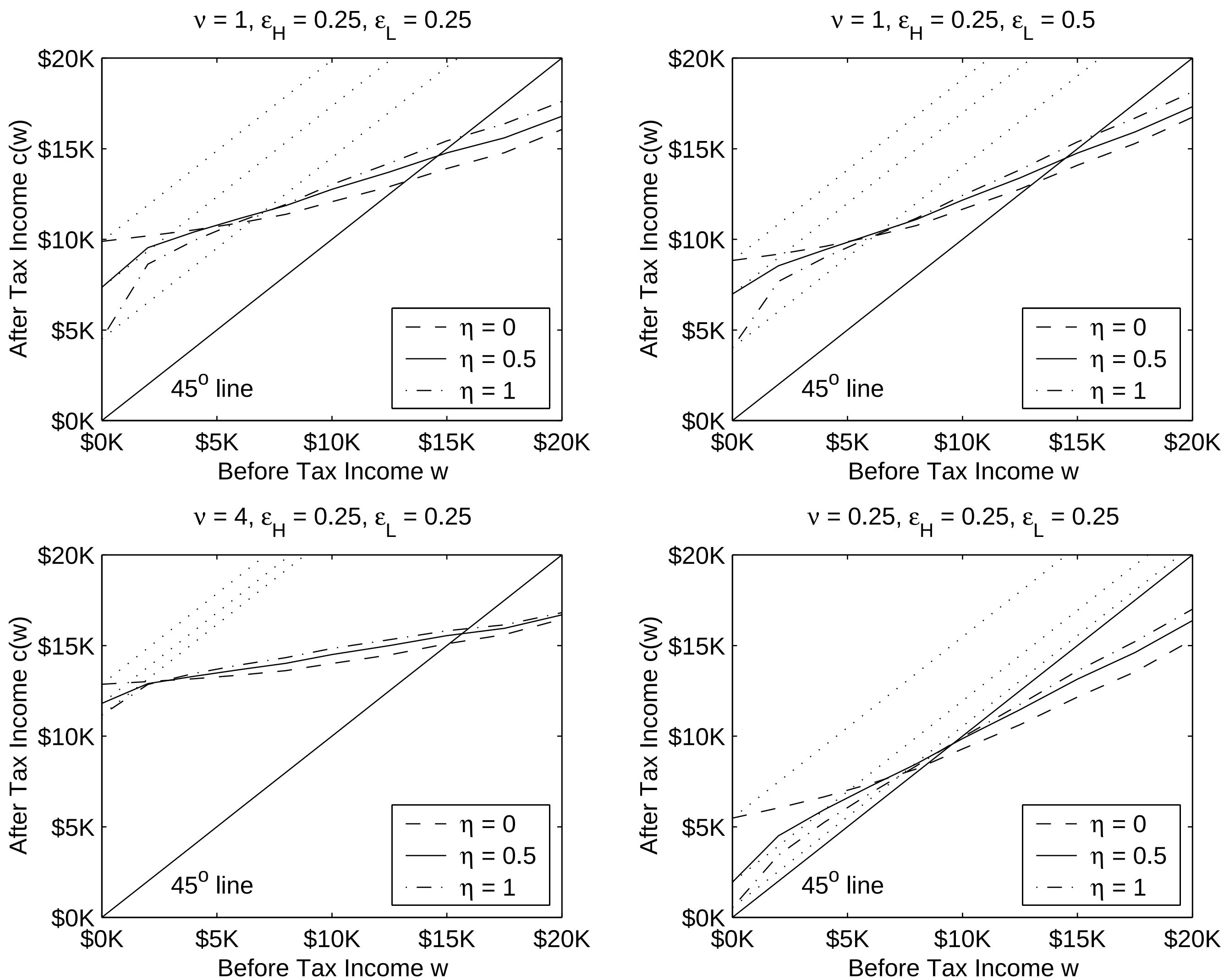

Table 1: Numerical Simulations with redistributive tastes parameter ν = 1.

	Middle-High Income Elasticity $\varepsilon_H = 0.25$						Middle-High Income Elasticity $\varepsilon_H = 0.5$					
	Guaranteed Income Level	Average M.T.Rate \$0-\$6K	Average M.T.Rate \$6K-\$15K	Break-Even Point	Average M.T.Rate \$30K+	Unem-ployment rate	Guaranteed Income Level	Average M.T.Rate \$0-\$6K	Average M.T.Rate \$6K-\$15K	Break-Even Point	Average M.T.Rate \$30K+	Unem-ployment rate
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Low Income Participation Elasticity η												
PANEL A: Intensive Elasticity for low incomes $e_L = 0$												
η = 0	$11,600	82	90	$14,100	42	15.6	$10,500	85	92	$10,700	31	15.7
η = 0.5	$7,700	7	84	$12,900	46	11.8	$6,900	15	87	$14,600	34	13.5
η = 1	$5,400	-23	71	$17,000	48	2.6	$4,300	-23	76	$13,800	35	2.9
PANEL B: Intensive Elasticity for low incomes $e_L = 0.25$												
η = 0	$9,900	83	67	$14,300	46	15.2	$8,800	85	69	$11,400	34	15.3
η = 0.5	$7,300	37	60	$12,900	49	13.8	$6,500	42	64	$10,300	36	14.2
η = 1	$4,500	-8	51	$16,800	50	2.5	$3,500	-7	55	$14,400	37	2.8
PANEL C: Intensive Elasticity for low incomes $e_L = 0.5$												
η = 0	$8,800	78	56	$14,400	48	15.0	$7,800	80	59	$11,800	35	15.0
η = 0.5	$7,000	45	50	$13,000	51	13.9	$6,200	51	54	$10,900	38	14.0
η = 1	$4,000	-2	42	$16,700	51	2.3	$3,000	0	45	$10,100	38	2.5

Notes: Simulations performed using redistributive taste parameter ν=1. No income effects included.

Table 2: Numerical Simulations with varying redistributive tastes

Low Income Participation Elasticity η	Elasticity $\varepsilon_H = 0.25$, Elasticity $\varepsilon_L = 0.25$					
	Guaranteed Income Level	Average M.T.Rate $0-$6K	Average M.T.Rate $6K-$15K	Break-Even Point	Average M.T.Rate $30K+	Unem-ployment rate
	(1)	(2)	(3)	(4)	(5)	(6)
	PANEL A: Low Redistributive tastes parameter ν = 0.25					
$\eta = 0$	$5,500	68	47	$9,600	27	14.6
$\eta = 0.5$	$1,900	12	34	$8,500	30	7.7
$\eta = 1$	$540	-5	25	$8,300	31	2.7
	PANEL B: Medium Redistributive tastes parameter ν = 1					
$\eta = 0$	$9,900	83	67	$14,300	46	15.2
$\eta = 0.5$	$7,300	37	60	$12,900	49	13.8
$\eta = 1$	$4,500	-8	51	$16,800	50	2.5
	PANEL C: High Redistributive tastes parameter ν = 4					
$\eta = 0$	$12,900	92	81	$17,400	62	16.0
$\eta = 0.5$	$11,800	69	79	$16,800	63	26.0
$\eta = 1$	$11,200	54	79	$16,600	64	27.2

Notes: Simulations performed using low income intensive elasticity $\varepsilon_L = 0.25$ and high income elasticity $\varepsilon_H = 0.25$.

Table A1: Empirical Earnings Distribution Calibration

Income Levels	Density weights (in percent)
(1)	(2)
$0	14.2
$2,000	3.3
$4,000	2.7
$6,000	2.8
$8,000	3.0
$10,000	4.8
$12,500	5.2
$15,000	6.5
$17,500	4.7
$20,000	8.2
$25,000	9.8
$30,000	16.4
$50,000	14.5
$100,000	3.9