

Essays on the Economics of Income Taxation

by

Emmanuel Saez

Submitted to the Department of Economics
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

May 1999

[June 1999]

© Massachusetts Institute of Technology 1999. All rights reserved.

Author

.....

Department of Economics

May 18, 1999

Certified by

.....

James Poterba

Professor of Economics

Thesis Supervisor

Certified by

.....

Peter Diamond

Professor of Economics

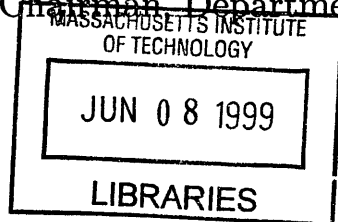
Thesis Supervisor

Accepted by

.....

Peter Temin

Chairman, Department Committee on Graduate Students



Essays on the Economics of Income Taxation

by

Emmanuel Saez

Submitted to the Department of Economics
on May 18, 1999, in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

Abstract

The first chapter derives optimal income tax formulas using the concepts of compensated and uncompensated elasticities of earnings with respect to tax rates. This method of derivation casts new light on the original Mirrlees formulas of optimal taxation and can be easily extended to a heterogeneous population of taxpayers. A simple formula for optimal marginal rates for high income earners is derived as a function of the two elasticities of earnings and the thickness of the income distribution. Optimal income tax simulations are presented using empirical wage income distributions and a range of realistic elasticity parameters.

The second chapter derives the non-linear income tax schedule which minimizes deadweight burden without any regard for redistribution. The features of this problem are shown to be equivalent to the Mirrlees' optimal income tax problem. The tax schedule minimizing deadweight burden is an optimal income tax schedule in which the government applies particular marginal welfare weights at each income level. In the case of no income effects, these marginal welfare weights are the same for everybody.

The last chapter uses a panel of individual tax returns and the 'bracket creep' as source of tax rate variation to construct instrumental variables estimates of the sensitivity of income to changes in tax rates. Compensated elasticities can be estimated by comparing the differences in changes in income between taxpayers close to the top-end of a tax bracket to the other taxpayers. The elasticities found are higher than those derived in labor supply studies but smaller than those found previously with the same kind of tax returns data.

Thesis Supervisor: James Poterba
Title: Professor of Economics

Thesis Supervisor: Peter Diamond
Title: Professor of Economics

Acknowledgments

I am very grateful to my advisors Peter Diamond and James Poterba for their guidance and support in my dissertation research. I am also very much indebted to Roger Guesnerie (my advisor at EHESS, France) and Thomas Piketty for their help and suggestions. I benefited from comments and stimulating discussions with Alberto Abadie, Josh Angrist, Abhijit Banerjee, Dan Feenberg, Xavier Gabaix, Jon Gruber, Michael Kremer and David Spector. Special thanks to Esther Duflo who read many drafts and provided me with very useful thoughts and suggestions.

I thankfully acknowledge financial support from the Department of Economics, MIT and the Alfred P. Sloan Foundation dissertation fellowship.

Contents

1	Using Elasticities to Derive Optimal Income Tax Rates	13
1.1	Introduction	13
1.2	Literature Review	16
1.3	Optimal Tax Rates: a Simple Approach	19
1.3.1	Elasticity concepts	19
1.3.2	High income optimal tax rates	21
1.3.3	Optimal Linear Rate	26
1.3.4	Conclusion	27
1.4	Optimal Tax Rates: General Results	28
1.4.1	The Mirrlees model	28
1.4.2	Optimal Marginal Rates	31
1.4.3	Optimal Asymptotic Rates	39
1.5	Empirical Results and Simulations	43
1.5.1	Optimal high income tax rates	43
1.5.2	Numerical simulations of tax schedules	48
1.6	Conclusion	52
2	A Characterization of the Income Tax Schedule Minimizing Deadweight Burden	81
2.1	Introduction	81
2.2	Deadweight Burden Concept	85
2.2.1	The Model	85
2.2.2	Deadweight Burden	86
2.3	Analysis with No Income Effects	87

2.3.1	Deadweight burden with no income effect	87
2.3.2	Minimization of DWB	88
2.3.3	Empirical Implementation	94
2.4	General Analysis	97
2.4.1	Deadweight Burden and Tax Reform	97
2.4.2	Links with Mirrlees Framework of Optimal Taxation	99
2.4.3	The limits of Consumer's Surplus concepts	102
2.5	Conclusion	104
3	The Effect of Marginal Tax Rates on Income:	
	A Panel Study of 'Bracket Creep'	119
3.1	Introduction	119
3.2	Literature	121
3.3	'Bracket Creep', Data and Descriptive Statistics	126
3.3.1	The 'Bracket Creep' phenomenon	126
3.3.2	Data	128
3.3.3	Descriptive Statistics	130
3.3.4	Wald Estimates	133
3.4	Model and Identification Strategy	135
3.5	Regression results	141
3.6	Conclusion	143

Introduction

This dissertation focuses on theoretical and empirical aspects of the economics of income taxation. There is a controversial debate about the degree of progressivity that the income tax should have. This debate is not limited to the economic research area but attracts much attention in the political sphere and among the public in general. At the center of the debate lies the equity-efficiency trade-off. Progressivity allows the government to redistribute from rich to poor because high incomes end up paying for a disproportionate share of public spending. But progressive taxation and high marginal tax rates have efficiency costs. High rates may affect the incentives to work and may therefore reduce the tax base (or even total tax receipts in the most extreme case), producing very large deadweight losses. The design of income tax schedules is one of the key elements of public policy.

Economists have devoted much effort to analyzing the income tax problem. The research on the subject is split into two areas. First, there is a large empirical literature that tries to estimate behavioral responses to taxation. The key parameters in this literature are the elasticities of income with respect to marginal tax rates. These elasticities are a measure of the size of the behavioral response and thus are good indicators of the efficiency costs of taxation.

Second, a theoretical literature has developed models to analyze the equity-efficiency trade-off: this is the theory of optimal income taxation. The modern framework for analyzing the equity-efficiency trade-off using a nonlinear income tax schedule was developed by Mirrlees (1971). It is intuitively clear that the size of the behavioral elasticities should critically affect optimal tax formulas. Therefore, optimal income tax theory should be closely related to the empirical literature on the behavioral responses to taxation. Unfortunately, this is not the case. The gap between the two literatures is in part the consequence of the theoretical approach that Mirrlees used to solve the problem. The theory of optimal taxation has been primarily used to derive general theoretical properties of tax schedules. However, the general results that have been derived are not strong enough to provide useful guidance for tax policy making. The literature has in general expressed optimal tax formulas in terms of the derivatives of utility functions, and in terms of an exogenous skill distribution, that

are hard to relate to empirical magnitudes. Very little has been done to relate the formulas to the elasticity concepts familiar in empirical studies or to investigate the link between the skill distribution and the income distribution.

The first chapter of the thesis derives optimal income tax formulas in terms of observable parameters, elasticities and shape of the income distribution, in order to bridge the gap between theory and applied studies. It finds that Mirrlees formulae of optimal income taxation can be derived directly using elasticities and that the optimal tax rate for high incomes can be expressed in a simple way in terms of elasticities and the thickness of the top income distribution tail.

The second chapter of the thesis derives the income tax schedule which minimizes deadweight burden with no regard for redistribution. The existing empirical literature on tax policy reforms usually relies on the concept of deadweight burden and uses Harberger's triangle approximation formula. This is not fully satisfying because Harberger's triangle formula is only an approximation valid in principle for small tax rates. My study shows that the non-linear income tax minimizing deadweight burden with no regard for redistribution is a particular Mirrlees's optimal income tax in which the government sets the same welfare weights for everybody. As a result, the shape of the income distribution and the size of income effects are important elements of the optimal tax problem with no regard for redistribution and thus Harberger's triangle formula fails to capture key elements of the income tax problem.

The last chapter contributes to the empirical literature on the behavioral responses to taxation. This study uses the 'bracket creep' in the federal income tax from 1979 to 1981 as a source of variation in tax rates to estimate the elasticity of reported income with respect to tax rates. The remainder of this introduction describes the three papers of the thesis in more detail.

The first chapter of the thesis revisits the Mirrlees (1971) theory of optimal income taxation in order to bridge the gap between that theory and the empirical literature on the behavioral responses to taxation. Instead of adopting the usual theoretical approach, my work starts directly from the behavioral elasticities familiar in empirical studies to derive optimal income tax results. More precisely, I consider small tax changes around the optimum tax schedule and analyse the effects of these

small changes on tax receipts and welfare. The effects due to behavioral responses are directly expressed in terms of elasticities. Around the optimal tax schedule, a small reform has no first order effects on welfare. This allows the derivation of first order conditions which are satisfied by the optimum tax schedule. Of course, these conditions are equivalent to the ones that Mirrlees obtained using direct mathematical methods of optimization. Nevertheless, my methodology, which uses elasticities directly, is fruitful for deriving results in optimal income taxation.

First, this method gives a clear understanding of the key factors underlying the general Mirrlees formula of optimal taxation. These are the behavioral elasticities (both substitution and income effects are important), the shape of the income distribution and the welfare weights that the government sets at each income level. More precisely, if the government increases the marginal tax rate locally around an income level z , all taxpayers with income above z will pay more taxes because total tax liability is simply the sum of marginal rates. However, because of behavioral responses, increasing tax rates around income level z reduces the incomes of taxpayers in that range. This negative effect is proportional to the average compensated elasticity and to the density of taxpayers around the income level z . Thus, the key element to determine the optimal marginal rate at income level z is the ratio of the number of taxpayers with incomes above z to the density of taxpayers at income level z . The shape of the income distribution is therefore of critical importance in the optimal tax formula even in the absence of redistributive motives. The formula I obtain displays clearly how efficiency and redistributive considerations interact in the optimal income tax problem.

Second, this chapter derives a simple formula for the optimal tax rate for high income taxpayers. The optimal top rate depends negatively on both uncompensated and compensated elasticities and positively on the thickness of the top tail of the income distribution. If the income distribution is bounded at the top, then the thickness of the top tail can be considered as equal to zero and the optimal top rate is also zero. This is the well known result of Seade and Sadka. Nevertheless, the empirical examination of wage income distributions shows that the thickness of the top tail of the distribution remains constant, and positive, over a very broad range of incomes. This thickness approaches zero only in the vicinity of the very

richest taxpayer. More precisely, the top tail of the income distribution is remarkably well approximated by a Pareto distribution whose tail is infinite. Thus, my formula, expressed in terms of elasticities and the Pareto parameter of the top tail (which measures the thickness of the tail), is valid over a very broad range of incomes and is therefore much more useful for practical fiscal policy purposes than the zero top result of Seade and Sadka. For example, if we assume that elasticities for high incomes are around 0.3 and that the government sets a relatively small social weight on the marginal consumption of high income earners, then the optimal top rate should be around 60%.

Third, my methodology using elasticities does not require the strong homogeneity assumptions of the Mirrlees model where all individuals have the same utility function and differ only through an exogenous skill parameter. The elasticity method can be extended much more easily to deal with a heterogeneous population of taxpayers. The same formulae remain valid once elasticities are considered as the average elasticities over the population at given income levels. The original Mirrlees' derivation relied heavily on the fact that all individuals shared the same utility function and differed only through their skill levels. That approach cannot be easily generalized to a heterogeneous population.

Last, because the formulas I have derived are closely related to empirical magnitudes such as elasticities and shape of the income distribution, I am able to present numerical simulations of optimal tax schedules based directly on the empirical income distribution. With constant elasticities, these simulations suggest that optimal marginal tax rates should be U-shaped. In the optimal income tax model, redistribution takes place through a negative income tax. Everybody is entitled to a guaranteed income level that is taxed away as earnings increase. The government applies very high rates at the bottom because it wants to tax away welfare quickly in order to target redistribution to low incomes. Marginal rates are increasing for high incomes because of the redistributive tastes of the government and because of the shape of the tail of the income distribution. The thickness of the top tail of the income distribution is measured by the ratio of the number of taxpayers above a given income level z to the density of taxpayers at income level z . This thickness parameter is increasing in z at the high end of income distribution.

An example illustrates the optimal tax rates simulations. If we assume that elasticities are constant and equal to 0.25 and that the government uses a redistributive utilitarian criterion, then tax rates at the bottom are around 75%. Tax rates decrease quickly until they reach a minimum of 40% at income level \$80,000 per year and per household. Tax rates then start to increase until they reach an asymptotic value of 60% at income level \$250,000. Above that income level, tax rates are roughly constant. The guaranteed income level is around one-third of the average income, or around \$15,000 per year.

The second chapter of the thesis considers the same income tax problem but abstracts from redistributive considerations and focuses on pure efficiency. This study solves for the income tax schedule which minimizes the excess burden and raises a given amount for public spending. As pointed out above, the deadweight burden has been the dominant concept in the applied literature when discussing tax policy. Most of the time, the well known Harberger's triangle approximation has been used. The triangle approximation has been an extremely useful tool because of its simplicity and its wide range of potential applications. The formula states that deadweight burden over the tax base is equal to one half of the compensated elasticity times the square of the marginal tax rate. The common wisdom that has emerged out of this approximation formula is that welfare loss depends on substitution elasticities and that the loss increases more than proportionally when the tax rate increases.

My study shows that the usual Harberger's deadweight burden approximation formula fails to capture key elements of the income tax problem. I show that solving for the tax schedule minimizing deadweight burden is in fact equivalent to solving the Mirrlees' optimal income tax problem. The tax schedule minimizing deadweight burden is an optimal income tax schedule in which the government applies particular marginal welfare weights at each income level. In the case of no income effects, these marginal welfare weights are the same for everybody. The tax schedule minimizing deadweight burden is then simply the Mirrlees optimal income tax with no redistributive concerns. At the optimum, the government is indifferent as whether it takes one dollar away from a wealthy taxpayer and transfers it a poorer taxpayer (or the reverse).

However, in the general case with income effects, the marginal welfare weights are

higher for people facing higher marginal rates. This property illustrates the deficiency of the deadweight burden concept in the presence of income effects. Most of the insights of optimal income taxation remain valid when minimizing deadweight burden with no regard to redistribution. In particular, the shape of income distribution has a strong impact on the pattern of marginal rates. This element is not captured by Harberger's triangle formula.

The third chapter of the thesis is devoted to the estimation of behavioral responses to taxation. This work is motivated by the observation that these parameters are crucial in the optimal tax formulas. However, empirical work has so far failed to generate a consensus on their values. It seems therefore important to think about why this is the case, and how elasticities could be consistently estimated. The classic labor supply literature has typically found small elasticities of hours of work with respect to tax rates. Subsequently, this view has been challenged by studies examining the effect of taxes on overall reported income and not just hours of work (Lindsey (1987), Feldstein (1995)). The ideal scheme for evaluating the impact of tax rates on income would be to increase tax rates randomly for some people and not others, and then compare their hours of work and income. In the absence of evidence from such an experiment, it is necessary to rely on legislated variations in tax rates. Researchers such as Feldstein (1995) and Lindsey (1987) have used the major tax reforms of the 1980's. They compared the growth rate in income of people affected by the reforms (high incomes) with that of people whose tax rates remained unchanged (low and middle incomes). This methodology amounts therefore to attributing the widening in inequalities to the tax reform. Economists have however proposed many non-tax explanations for the increase in inequalities during that period.

In this paper, I use the bracket creep in the federal income tax from 1979 to 1981 as source of variation in tax rates. During that period, inflation was still high (around 10% per year), while the tax schedule was fixed in nominal terms. Because the income tax schedule was highly progressive, inflation had a strong real impact. The income levels at which marginal rates change shifted down in real terms. Therefore a taxpayer near the end of a bracket was likely to *creep* to the next bracket even if his income did not change in real terms, while a taxpayer far from the end of a bracket was not as likely to experience a rise in marginal rate the following year. As a result, people

whose incomes were very similar, experienced a very different evolution in marginal rates. The spirit of the identification strategy is to compare the variations in income among people who are near the end of a bracket and people who are far from it. For a given bracket, people have in fact very similar incomes. Therefore it is possible to consider that the variation in tax rates faced by the affected group is exogenous, after I control flexibly for income in the regression.

The changes in marginal rates induced by “bracket creep” are small (from 4 to 7 percentage points) compared to the drastic tax cuts of the 1980s. However, for most of the 15 brackets, I find that taxpayers who are affected by “bracket creep” experience smaller income growth than comparable but non affected individuals. This suggests that taxpayers do indeed react to taxes. The implied elasticities range between 0.2 and 0.5 for taxable income, 0.1 and 0.4 for adjusted gross income. These elasticities are substantially smaller than the ones found in the studies using tax reforms. Elasticities for wage income are never significant and close to zero which is consistent with the labor supply literature. The discrepancy between the estimates for taxable income, adjusted gross income and wages suggests that most of the response is a consequence of reporting behavior rather than of a change in actual labor supply.

This thesis tries to cast light on the income tax problem by studying both the theoretical and empirical aspects of the problem. The main contribution of this work is to describe optimal income tax results in the light of the elasticity concepts that are used in applied studies. This methodology considerably improves our understanding of the non-linear income tax problem.

The present work could be extended in several directions. First, my theoretical work has derived important parameters related to the shape of income distribution which have important consequences for the optimal tax problem. I have estimated these parameters only for the US and for a short period of time. It would be interesting to extend this study to other countries and other time periods in order to see whether the results obtained here are valid in other contexts. Second, it might be fruitful to apply the same methodology to other tax and redistribution problems. In particular, the issue of optimal tax rates at the bottom of income distribution deserves more attention in order to cast light on the important problem of the design of income maintenance programs.

Chapter 1

Using Elasticities to Derive Optimal Income Tax Rates

1.1 Introduction

There is a controversial debate about the degree of progressivity that the income tax should have. This debate is not limited to the economic research area but attracts much attention in the political sphere and among the public in general. At the center of the debate lies the equity-efficiency trade-off. Progressivity allows the government to redistribute from rich to poor because high incomes end up paying for a disproportionate share of public spending. But progressive taxation and high marginal tax rates have efficiency costs. High rates may affect the incentives to work and may therefore reduce the tax base (or even total tax receipts in the most extreme case), producing very large deadweight losses. The modern setup for analyzing the equity-efficiency tradeoff using a general nonlinear income tax was built by Mirrlees (1971). Since then, the theory of optimal income taxation based on the original Mirrlees's framework has been considerably developed. The implications for policy, however, are limited for two main reasons. First, optimal income tax schedules have few general properties: we know that optimal rates must lie between 0 and 1 and that they equal zero at the top and the bottom. These properties are of little practical relevance for tax policy. In particular the zero marginal rate at the top is a very local result which applies only at the very top and is not robust when uncertainty is introduced

in the model; it is therefore of no practical interest. Moreover, numerical simulations tend to show that tax schedules are very sensitive to utility functions chosen (see for example Tuomala (1990), Chapter 6).

Second, optimal income taxation has interested mostly theorists and has not changed the way applied public finance economists think about the equity-efficiency tradeoff. Theorists are mostly interested in general qualitative properties of utility functions and tax schedules whereas elasticities are the key concept in applied studies. There has been no systematic attempt to derive results in optimal taxation which could be easily used in applied studies. Most of the empirical literature on the behavioral effects of income taxation tries to estimate elasticities of income (such as wage income, capital gains or overall taxable income) with respect to marginal rates. Once elasticities are computed, optimal taxation theory is often ignored and tax reform discussions are centered on the concept of deadweight burden.¹ Therefore, most discussions of tax reforms focus only on the costs of taxation but are unable to weight both costs and benefits to decide whether taxes are too high or too low.

This paper argues that there is a simple link between optimal tax formulas and elasticities of income familiar to empirical studies. The aim of an optimal income tax (in addition to meeting government's revenue needs) is to redistribute income to the poor. The income tax, however, produces distortions and may have negative effects on labor supply and thus can reduce income and even total taxes collected. Therefore, what is important to know is whether the wealthy continue to work when tax rates increase (without utility compensation); the uncompensated elasticity² is thus likely to play a bigger role than the compensated elasticity in optimal tax formulas. In other words, this paper shows that the precise division of compensated effects into uncompensated effects and income effects plays a major role in optimal taxation. However, the empirical literature has rarely paid much attention to this division because it focused almost exclusively on deadweight burden approximations.

Recently, Diamond (1998) has taken an important step toward the narrowing of

¹The deadweight burden is a measure of the inefficiency of taxation. The approximation commonly used, known as Harbarger's triangle formula, is proportional to the compensated elasticity of income with respect to marginal tax rates.

²The uncompensated elasticity is equal to the compensated elasticity minus revenue effects by the Slutsky equation. See Section 3.1.

the gap between optimal taxation theory and practical policy recommendations by considering quasi-linear utility functions and analyzing precisely the influence of elasticities of labor supply and the shape of the wage rate distribution on the optimal tax schedule. Using quasi-linear utility functions is equivalent to assuming no income effects and thus Diamond (1998) could not examine the role of income effects. It turns out that his results can be considerably generalized and that a very simple formula for high income tax rates can be derived in terms of both the compensated and uncompensated tax rate elasticities of incomes and the thickness of the top tail of the income distribution. Expressing the first order condition for optimal rates in terms of elasticities simplifies considerably the general Mirrlees formula and gives a much better understanding of the key economic effects that underlie it. Moreover, the optimal tax formulas derived using elasticities can be easily extended to a heterogeneous population.

Empirical studies provide a wide range of elasticity estimates but the thickness of the tail of the income distribution has not been studied extensively for practical taxation purposes because it does not enter the deadweight burden approximation formula and thus has not been considered as a crucial element when discussing tax policy. This paper also examines the empirical distributions of earned income using tax returns data and displays simulations of optimal income tax schedules using empirical distributions of income and making realistic assumptions about elasticity parameters.

The paper is organized as follows. Section 2 reviews the main results of the optimal income tax literature. Section 3 first recalls the usual results about elasticities of earnings. It then derives a simple formula for optimal high income tax rates. The optimal linear income tax is also examined. Section 4 presents the theoretical results of this paper in the framework of the Mirrlees model. The general Mirrlees first order condition for optimal rates is reexamined in terms of elasticities. The relation between the distribution of skills and the distribution of incomes is examined and optimal asymptotic tax rates are derived. Section 5 discusses the elasticity results of the empirical taxation literature and presents empirical results about wage income distributions along with numerical simulations of optimal tax rates. Section 6 concludes and discusses policy implications. The main results of this paper can be understood

without relying explicitly on the Mirrlees framework of optimal income taxation. Section 4 is more technical but can be skipped without affecting the understanding of the subsequent sections.

1.2 Literature Review

The Mirrlees framework captures the most important features of the tax design problem. The economy is competitive and households differ only in the levels of skills in employment. Households supply labor elastically and thus taxation has efficiency costs. The government wants to maximize a social welfare function but cannot observe skills; it must therefore rely on a distortionary nonlinear income tax to meet both its revenue requirements and redistribute income.

General results about optimal tax schedules are fairly limited. Tuomala (1990) (Chapter 6) and Myles (1995) (Chapter 5) present most of the formal results. Mirrlees (1971) showed that there is no gain from having marginal tax rates above 100 percent because nobody will choose to have such a rate at the margin. Under reasonable assumptions for the utility function, optimal marginal rates cannot be negative either. Mirrlees (1971) presented these properties and Seade (1982) clarified the conditions under which they hold.

The most striking and well known result is that the marginal tax rate should be zero at the income level of the top skill if the distribution of skills is bounded (Sadka (1976) and Seade (1977)). The argument for this result is intuitive: if the rate faced by the top earner is larger than zero, then, extending the tax schedule to higher incomes with a zero tax rate would lead the top earner to work more and would not reduce tax revenue and thus would lead to a Pareto improvement. Numerical simulations (see for example Tuomala (1990)) have shown, however, that this result is very local. Optimal rates do not approach zero until very close to the top and thus this result is of little practical interest. Mirrlees (1971) did not derive this simple result because he considered unbounded distributions of skills. He nonetheless presented precise conjectures about asymptotic optimal rates in the case of utility functions separable in consumption and labor (Mirrlees (1971), p.189). The optimal asymptotic formulas he derived were simple; they showed clearly that optimal asymptotic rates depend positively on

the thickness of the tail of the skill distribution. Nonetheless, these conjectures have remained practically unnoticed in the subsequent optimal income tax literature. This can be explained by two reasons. First, Mirrlees conjectures depend not only on the distribution of skills (which is already unobservable empirically) but also on abstract properties of the utility function with no obvious intuitive meaning. Second the zero top rate result was probably considered for a long time as the definitive result because commonsense would suggest that a finite distribution of skills is closer to the reality than an unbounded one. This paper generalizes and gives a simple interpretation of the early Mirrlees conjectures. Moreover, the empirical results will show that in fact unbounded distributions are of much more interest than bounded distributions to approximate optimal tax rates for high income earners.

In addition to the zero top result, a few more results have been derived for the bottom of the skill distribution. If everybody works (and supplies labor bounded away from zero) then Seade (1977) showed that the bottom rate is also zero. However, if there is an atom of non workers then the bottom tax rate is positive (Ebert (1992)). This later case is probably the most relevant empirically.

Recently, Atkinson (1990) using quasi-linear utility functions with constant labor supply elasticity noticed that the top rate converges to a simple limit when the skill distribution is Pareto distributed. Diamond (1998) extended this particular case and began to examine empirical distributions. Moreover, he obtained simple results about the pattern of the marginal rates as a function of simple properties of the distribution of skills.

Piketty (1997) considered the same quasi-linear utility case and derived Diamond's optimal tax formulas for the Rawlsian criterion without setting a formal program of maximization. He considered instead small local changes in marginal rates and used directly the elasticity of labor supply to derive the behavioral effects of this small reform. The optimal rate can be derived using the fact that at the optimum, the small tax reform should lead to zero first order effect on tax receipts. My paper clarifies and generalizes this alternative method of derivation of optimal taxes.³

Another strand of the public economics literature has developed similar elasticity

³I am indebted to Thomas Piketty for his suggestions and help in deriving my results using this alternative method.

methods to calculate the marginal costs of public funds. The main purpose of this literature was to develop tools more sophisticated than simple deadweight burden computations to evaluate the efficiency costs of different kinds of tax reforms and the optimal provision of public goods (see for example Mayshar (1991), Ballard and Fullerton (1992) and Dahlby (1998)). Because this literature was mainly interested in assessing the efficiency of existing tax schedules and not in computing optimal tax schedules, the links between this literature and the optimal income tax literature have been very limited. I will show that the methods of this literature can be useful to derive results in optimal taxation and that, in particular, Dahlby (1998) has come close to my results for high income rates.

Starting with Mirrlees (1971), considerable effort has gone into simulations of optimal tax schedules. Following Stern (1976), attention has been paid on a careful calibration of the elasticity of labor supply. Most simulation results are surveyed in Tuomala (1990). It has been noticed that the level of inequality of the distribution of skills and the elasticities of labor supply⁴ significantly affect optimal schedules. Nevertheless, simulations did not lead researchers to conjecture or prove a general result for top rates because most simulations use a log-normal distribution of skills which matches roughly the single moded empirical distribution but has also an unrealistically thin top tail and leads to marginal rates converging to zero (Mirrlees (1971)).

Nobody has tried to use empirical distributions of income to perform simulations because the link between skills and realized incomes was never investigated in depth. This study shows that for high income earners, a simple relation can be derived between the distribution of skills and the distribution of incomes. As a result, it is possible to use empirical distributions of income to perform simulations of optimal tax rates which may provide useful practical policy recommendations.

⁴The numerical simulations focus on the elasticity of substitution between labor and consumption instead of uncompensated and compensated elasticities of labor supply.

1.3 Optimal Tax Rates: a Simple Approach

The aim of this Section is to show that the familiar concepts of compensated and uncompensated elasticities of earnings with respect to marginal tax rates can be useful to derive in a simple way interesting results about optimal tax rates. I first consider the problem of the optimal tax rate for high income earners and then the problem of the optimal linear tax.

To deal with the first problem, I consider that the government sets a flat marginal rate τ above a given (high) income level \bar{z} and then I consider the effects of a small increase in τ on tax receipts for the government and on social welfare. The behavioral responses can be easily derived using the elasticities. The government sets the optimal tax rate τ such that a small increase in tax rates has no first order effects on total social welfare.⁵

The problem of the optimal linear tax can be solved in a similar way by considering small increases in the optimal flat rate and in the lump sum amount redistributed to every taxpayer. Before presenting the results, I recall the definitions of the elasticities which are used throughout the paper.

1.3.1 Elasticity concepts

I consider a standard two good model. A taxpayer maximizes an individual utility function $u = u(c, z)$ which depends positively on consumption c and negatively on earnings z . The utility function represents strictly convex preferences. This framework is a simple extension of the standard labor supply model where utility depends on consumption and labor supply and where earnings is equal to labor supply times an exogenous pre-tax wage rate.⁶ Assuming that the individual is on a linear portion of the tax schedule, the budget constraint can be written as $c = z(1 - \tau) + R$, where

⁵Dahlby (1998) considered piecewise linear tax schedules and used the same kind of methodology to compute the effects of a general tax rate reform on taxes paid by a “representative” individual in each tax bracket. By specializing his results to a reform affecting only the tax rate of the top bracket, he derived a formula for the tax rate maximizing taxes paid by the “representative” individual of the top bracket. In this Section, I study carefully the issue of aggregation across individuals and show how this method can lead to interesting optimal tax rate results.

⁶My formulation is more general because it allows for potential endogeneity between the wage rate and labor supply.

τ is the marginal tax rate and R is defined as virtual income. Virtual income is the post-tax income that the individual would get if his earnings were equal to zero was allowed to stay on the “virtual” linear schedule. The first order condition of the individual maximization program, $(1 - \tau)u_c + u_z = 0$, defines implicitly a Marshallian (uncompensated) earnings supply function $z = z(1 - \tau, R)$ which depends on (one minus) the marginal tax rate τ and the virtual income R . From this earnings supply function, the usual concepts of elasticity of earnings and marginal propensity to earn out of non wage income⁷ can be defined. The uncompensated elasticity (denoted by ζ^u) is defined such that:

$$\zeta^u = \frac{1 - \tau}{z} \frac{\partial z}{\partial(1 - \tau)} \quad (1.1)$$

The marginal propensity to earn out of non wage income (denoted by mpe) is such that:

$$mpe = (1 - \tau) \frac{\partial z}{\partial R} \quad (1.2)$$

The Hicksian (compensated) earnings function can be defined as the earnings level which minimizes cost $c - z$ needed to reach a given utility level u for a given tax rate τ . I denote it by $z^c = z^c(1 - \tau, u)$. The compensated elasticity of earnings ζ^c is defined by:

$$\zeta^c = \frac{1 - \tau}{z} \frac{\partial z}{\partial(1 - \tau)} \Big|_u \quad (1.3)$$

The two elasticity concepts and the revenue effects are related by the Slutsky equation:

$$\zeta^c = \zeta^u - mpe \quad (1.4)$$

The compensated elasticity is always non-negative and mpe is non positive if leisure is not an inferior good, an assumption I make from now on. The sign of the uncompensated elasticity is ambiguous but the uncompensated elasticity is always smaller (or equal) than the compensated elasticity. Note that these definitions are identical to usual definitions of elasticities of labor supply if one assumes that the

⁷See Pencavel (1986) for a more detailed presentation.

wage rate w is exogenous and that earnings z is equal to wl where l represent hours of work.

1.3.2 High income optimal tax rates

I assume in this subsection that the government wants to set a constant linear rate τ of taxation above a given (high) level of income \bar{z} . I normalize without loss of generality the population with income above \bar{z} to one and I denote by $h(z)$ the density of the income distribution. The goal of this subsection is to find out the optimal τ for the government.

I consider a small increase $d\tau$ in the top tax rate τ for incomes above \bar{z} . Clearly, this tax change does not affect taxpayers with income below \bar{z} . The tax change can be decomposed into two parts (see Figure 1); first, an overall uncompensated increase $d\tau$ in marginal rates (starting from 0 and not just from \bar{z}), second, an overall increase in virtual income $dR = \bar{z}d\tau$. For a given individual earning income z (above \bar{z}), total taxes paid are equal to $T(z) = \tau[z(1 - \tau, R) - \bar{z}] + T(\bar{z})$. The small tax reform produces the following effect on his tax liability:

$$\frac{\partial T(z)}{\partial \tau} = (z - \bar{z}) + \tau \left[-\frac{\partial z}{\partial (1 - \tau)} + \frac{\partial z}{\partial R} \bar{z} \right] \quad (1.5)$$

Therefore, this tax change has two effects on tax liability. First, there is a mechanical effect (first term in parentheses in equation (1.5)) and second, there is a behavioral effect (second term in square brackets in equation (1.5)). Let us examine these two effects successively.

- **Mechanical effect**

The mechanical effect (denoted by M) represents the increase in tax receipts if there were no behavioral responses. A taxpayer with income z (above \bar{z}) would pay $(z - \bar{z})d\tau$ additional taxes. This is the first term in equation (1.5). Therefore, summing over the population above \bar{z} and denoting the mean of incomes above \bar{z} by z_m , the total mechanical effect M is equal to,

$$M = [z_m - \bar{z}]d\tau \quad (1.6)$$

• **Behavioral Response**

The behavioral response effect (denoted by B) can in turn be decomposed into the two effects displayed in Figure 1: first, an uncompensated elastic effect (first term in the square bracket expression in equation (1.5)) and second, an income effect (second term in the square bracket expression in equation (1.5)). The uncompensated effect is the behavioral response of taxpayers to the increase in tax rate $d\tau$. By definition of the uncompensated elasticity, the response of a taxpayer earning z is equal to $-\zeta^u z d\tau / (1 - \tau)$. The income effect is due to the increase in virtual income $dR = \bar{z} d\tau$. By definition of mpe , the response of an individual earning z is equal to $mpe \bar{z} d\tau / (1 - \tau)$. The total behavioral response dz of an individual is the sum of these two effects:

$$dz = -(\zeta^u z - mpe \bar{z}) \frac{d\tau}{1 - \tau} \quad (1.7)$$

If $z \gg \bar{z}$ the income effect component is negligible and the response is fully uncompensated. If $z \simeq \bar{z}$, then, using the Slutsky equation (1.4), the response is approximately equal to $-\zeta^c z d\tau / (1 - \tau)$; the response is therefore fully compensated. Equation (1.7) is important to bear in mind when tax reforms are used to estimate elasticities.⁸

The reduction in income dz displayed in equation (1.7) implies a reduction in tax receipts equal to τdz . The total reduction in tax receipts due to the behavioral responses is simply the sum of the terms τdz over all individuals earning more than z ,

$$B = -\bar{\zeta}^u z_m \frac{\tau d\tau}{1 - \tau} + \overline{mpe} \frac{\bar{z} \tau d\tau}{1 - \tau} \quad (1.8)$$

where $\bar{\zeta}^u = \int_{\bar{z}}^{\infty} \zeta_{(z)}^u z h(z) dz / z_m$ is a weighted average of the uncompensated elasticity. The elasticity term $\zeta_{(z)}^u$ inside the integral represents the average elasticity over individuals earning income z . $\overline{mpe} = \int_{\bar{z}}^{\infty} mpe_{(z)} h(z) dz$ is the average of $mpe_{(z)}$.⁹ Note that \overline{mpe} and $\bar{\zeta}^u$ are not averaged with the same weights. It is not necessary to

⁸In particular, if a tax reform adds a bracket at income level \bar{z} , comparing the responses of taxpayers just below \bar{z} and just above \bar{z} allows a simple estimation of ζ^c .

⁹ $mpe_{(z)}$ is the average income effect for individuals earning z .

assume that people earning the same income have the same elasticity; the relevant parameters are simply the average elasticities at given income levels.

Adding equations (1.6) and (1.8), the overall effect of the tax reform on government's revenue is obtained,

$$M + B = \left[\frac{z_m}{\bar{z}} - 1 - \frac{\tau}{1 - \tau} (\bar{\zeta}^u \frac{z_m}{\bar{z}} - \overline{mpe}) \right] \bar{z} d\tau \quad (1.9)$$

The tax reform raises revenue if and only if the expression in square brackets is positive. If the government values much more an additional dollar given to the poorest people than the same additional dollar given to the top bracket taxpayers,¹⁰ then it will raise the maximum amount of taxes from the top bracket taxpayers. In that case, it will set the top rate τ such that the expression in square brackets in equation (1.9) is equal to zero.

In the general case, let us consider \bar{g} which is the ratio of social marginal utility for the top bracket taxpayers to the marginal value of public funds for the government. In other words, \bar{g} is defined such that the government is indifferent between \bar{g} more dollars of public funds and one more dollar consumed by the taxpayers with income above \bar{z} . \bar{g} can be considered as a parameter reflecting the redistributive goals of the government.

Each additional dollar raised by the government because of the tax reform reduces social welfare of people in the top bracket by \bar{g} and thus is valued only $1 - \bar{g}$ by the government. First order behavioral changes in earnings lead only to second order effects on welfare (this is the usual consequence of the envelope theorem). As a result, the loss of one dollar in taxes due to behavioral effects is valued one dollar (and not $1 - \bar{g}$ dollars) by the government. Consequently, the government wants to set the rate τ such that, $(1 - \bar{g})M + B = 0$. Thus the optimal rate is such that,

$$\frac{\tau}{1 - \tau} = \frac{(1 - \bar{g})(z_m/\bar{z} - 1)}{\bar{\zeta}^u z_m/\bar{z} - \overline{mpe}} \quad (1.10)$$

which leads to,

¹⁰This is of course the case with the Rawlsian criterion. This is also the case with a utilitarian criterion if one considers utility functions with marginal utility of consumption declining to zero as consumption tends to infinity.

$$\tau = \frac{1 - \bar{g}}{1 - \bar{g} + \bar{\zeta}^u(\frac{z_m}{\bar{z}})/(\frac{z_m}{\bar{z}} - 1) - \overline{mpe}/(\frac{z_m}{\bar{z}} - 1)} \quad (1.11)$$

This equation gives a strikingly simple answer to the problem of the optimal marginal rate for high income earners. This formula applies to heterogeneous populations. The relevant parameters are the weighted average elasticities $\bar{\zeta}^u$ and average income effects \overline{mpe} which can be estimated empirically. The optimal rate τ is a decreasing function of \bar{g} , $\bar{\zeta}^u$ and $-\overline{mpe}$ (absolute size of income effects) and an increasing function of z_m/\bar{z} .

The ratio z_m/\bar{z} of the mean of incomes above \bar{z} to the income level \bar{z} , is larger than one. From now on, I call this ratio the conditional mean income ratio. If the tail of the income distribution follows a Pareto distribution with parameter $a > 1$ ($Prob(Income > z) = C/z^a$) then the density of incomes $h(z)$ is then to aC/z^{1+a} . In that case, it is easy to show that z_m/\bar{z} is constant and equal to $a/(a-1)$.¹¹

If the tail of the distribution is thinner than any Pareto distribution (e.g., such as a log-normal or an exponential distribution) then z_m/\bar{z} tends to 1 and therefore we can consider that in this case $a = \infty$. Section 5 will show that, empirically, z_m/\bar{z} is strikingly stable over a very large range of incomes. Therefore, the tails of empirical earnings distributions can be remarkably well approximated by Pareto distributions.¹² The parameter a is approximately equal to 2.

Let me now consider the asymptotics of equation (1.11). Assuming that z_m/\bar{z} converges to a value (say m_∞), I can define a (between one and infinity) such that $a/(a-1) = m_\infty$. a can be considered as the limiting ‘‘Pareto’’ parameter of the income distribution. If $\zeta_{(z)}^u$ and $mpe_{(z)}$ converge to limiting values (denoted also by $\bar{\zeta}^u$ and \overline{mpe}) when z tends to infinity, the Slutsky equation¹³ implies that $\zeta_{(z)}^c$ converges to $\bar{\zeta}^c$ such that $\overline{mpe} = \bar{\zeta}^u - \bar{\zeta}^c$. In this case, (1.11) can be rewritten as a function of a and the limiting values of the elasticities $\bar{\zeta}^c$ and $\bar{\zeta}^u$:

¹¹ a must be larger than one to rule out infinite aggregate income $\int zh(z)dz$.

¹²This is of course not a new finding. Pareto discovered this empirical regularity more than a century ago (see Pareto (1965)). That is why these power law densities are called Pareto distributions.

¹³It is not possible to use directly the Slutsky equation in (1.11) because ζ^u and mpe are not averaged with the same weights.

$$\bar{\tau} = \frac{1 - \bar{g}}{1 - \bar{g} + \bar{\zeta}^u + \bar{\zeta}^c(a - 1)} \quad (1.12)$$

When these parameters do converge, the government wants to set roughly the same linear rate $\bar{\tau}$ above any large income level and thus $\bar{\tau}$ is indeed the optimal non-linear asymptotic rate of the Mirrlees problem.¹⁴ I show in Section 4 that the parameter a is independent of $\bar{\tau}$ as long as $\bar{\tau} < 1$. The intuition is the following: when elasticities are constant, changing the tax rate has the same multiplicative effect on the incomes of each high income taxpayer and therefore the ratio z_m/\bar{z} is unchanged. Empirically, a does not seem to vary with level of the top rate. I come back to this point in Section 5 but a thorough empirical investigation of this issue is left for future research.

$\bar{\tau}$ is decreasing in the four parameters $\bar{\zeta}^c$, $\bar{\zeta}^u$, a and \bar{g} . This is hardly surprising. Interestingly, for a given compensated elasticity $\bar{\zeta}^c$, the precise division into income effects and uncompensated rate effects matters. The higher are absolute income effects ($-\bar{m}\bar{p}\bar{e}$) relative to uncompensated effects ($\bar{\zeta}^u$), the higher is the asymptotic tax rate $\bar{\tau}$. This result confirms the intuition developed in the introduction: what matters most for optimal taxation is whether taxpayers continue to work when tax rates increase (without utility compensation).

The top rate $\bar{\tau}$ also depends negatively of the thickness of the top tail distribution measured by the Pareto parameter a or the limiting value of z_m/\bar{z} . This is also an intuitive result: if the distribution is thin then raising the top rate for high income earners will raise little additional revenue because the mechanical effect M depends on the difference between z_m and \bar{z} while the distortions are proportional to z_m (for the uncompensated effect) and \bar{z} (for the income effect) and thus are high at high income levels. If the distribution of income is bounded, then close to the top, \bar{z} is close to z_m and so the conditional mean income ratio tends to one and thus the top rate must be equal to zero (see equation (1.10)). This is the classical zero top rate result derived by Sadka (1976) and Seade (1977). If the tail is infinite but thinner than any Pareto distribution (i.e., $a = \infty$) then the asymptotic rate must also be zero.

¹⁴This point is proved rigorously in the Mirrlees model in Section 4.

1.3.3 Optimal Linear Rate

The analysis above can be easily applied to the case of the optimal linear tax rate. Many papers (beginning with Sheshinski (1972)) have studied this case but no paper has derived the optimal rate using directly the concepts of elasticities of earnings and marginal propensity to earn out of non wage income.¹⁵

In the case of an optimal linear tax, the government imposes a budget constraint of the form: $c = (1 - \tau)z + R$ by choosing the tax rate τ and a lump-sum level R . I note $H(z)$ the distribution of income and $h(z)$ its density function. I note $MS(z)$ the social marginal utility of consumption for individuals with income z and p the social value of public funds.

Consider first an increase of the tax rate from τ to $\tau + d\tau$, then, an individual with income z will pay $zd\tau$ additional taxes (mechanical effect), valued only $z(1 - MS(z)/p)d\tau$ by the government. Moreover, the individual will change its earnings by $dz = -\zeta^u z d\tau / (1 - \tau)$ which changes the amount of taxes it pays by τdz . The effect aggregated over the population must be null at the optimum and therefore:

$$1 - \int \frac{MS(z)}{p} \frac{z}{z_M} h(z) dz = \frac{\tau}{1 - \tau} \bar{\zeta}^u \quad (1.13)$$

where $z_M = \int zh(z)dz$ denotes average income and $\bar{\zeta}^u = \int \zeta^u zh(z)dz / z_M$ is a weighted average of the uncompensated elasticity.

Next, suppose that the government increases the lump sum R by dR , then the tax collected on a given individual earning z decreases by dR but the social loss is only $(1 - MS(z)/p)dR$. The individual changes its earnings by $dz = mpe dR / (1 - \tau)$ which changes the amount of taxes it pays by τdz . The overall aggregated effect must be null at the optimum and thus,

$$1 - \frac{\overline{MS}}{p} = \frac{\tau}{1 - \tau} \overline{mpe} \quad (1.14)$$

where $\overline{MS} = \int MS(z)h(z)dz$ denotes average social marginal utility and $\overline{mpe} = \int mpe h(z)dz$ is the average of mpe .

¹⁵The only exception is Piketty (1997) who derived the optimal linear rate for the simple case of the Rawlsian criterion with the method used here.

Equations (1.13) and (1.14) can be combined to eliminate p , and to obtain the following formula for the optimal tax rate τ^* :

$$\tau^* = \frac{1 - \bar{G}}{1 - \bar{G} + \bar{\zeta}^u - \bar{G} \bar{mpe}} \quad (1.15)$$

where \bar{G} is defined such that,

$$\bar{G} = \int \frac{MS(z)}{\bar{MS}} \frac{z}{z_M} h(z) dz$$

If $MS(z)$ is decreasing (which is a reasonable assumption if the government has redistributive goals), then $\bar{G} < 1$.¹⁶ \bar{G} is the smaller, the greater is inequality and the greater are the redistributive goals of the government. In the Rawlsian case, $MS(z) = 0$ for every z positive which implies $\bar{G} = 0$. Using equation (1.13), we obtain $\tau^* = 1/(1 + \bar{\zeta}^u)$. In any case, \bar{G} can be considered as a parameter chosen by the government according to its preferences. Once a distribution of incomes $H(z)$ is given, the government chooses the function $MS(z)$ and thus can compute \bar{G} .¹⁷ Unsurprisingly, the optimal linear rate is decreasing in \bar{G} , in the size of the uncompensated elasticity and in the absolute size of income effects.

These results can be derived in the classical model of optimal linear taxation. The interpretations of the optimal rate formula are often close to the one presented here (see for example Atkinson-Stiglitz (1980), pp. 407-408). However, I presented the results without referring to a distribution of skills to show that formula (1.15) can be applied in a much more general framework with heterogeneous agents. Similarly to the previous subsection, the only thing that matters is average elasticities; these average elasticities can be measured empirically.

1.3.4 Conclusion

This Section has shown that considering small reforms around the optimum and deriving the behavioral responses using elasticity concepts is a natural way to derive

¹⁶This is because $\bar{G} \bar{MS}$ is an average of $MS(z)$ with weights $zh(z)/z_M$ which overweighs high z and thus $\bar{G} \bar{MS}$ is smaller than \bar{MS} .

¹⁷Note however that in the optimal linear tax case, \bar{G} is positive except in the extreme Rawlsian case whereas in the asymptotic non-linear tax case of Section 3.2, \bar{g} could be zero even with a utilitarian criterion.

optimal tax rate results. Formulas for optimal rates (1.12) and (1.15) show that the pattern of elasticities as well as the shape of the income distribution and the redistributive goals of the government are the relevant parameters. In particular, though ζ^c is a sufficient statistic to approximate the deadweight loss of taxation, same values of ζ^c can lead to very different optimal tax rates. The bigger the income effects relative to uncompensated elasticity, the higher is the optimal tax rate.

In the next Section, I show that the systematic use of elasticity concepts in the general Mirrlees model is fruitful. First, by considering as in this Section a small tax reform (a small local increase in marginal rates), it is possible to derive the general Mirrlees formula for optimal tax rates without referring to adverse selection theory. This derivation allows a better grasp on the different effects at play than blind mathematical optimization and can be easily extended to heterogeneous populations. Second, it will be shown that the income distribution and the skill distribution are closely related through the uncompensated elasticity. This result is of crucial importance to perform numerical simulations (presented in Section 5) using empirical earnings distributions.

1.4 Optimal Tax Rates: General Results

1.4.1 The Mirrlees model

In the model, all individuals have identical preferences. The utility function depends on composite consumption c and labor l and is noted $u(c, l)$. I assume that preferences are well behaved and that u is regular (at least of class C^4). The individuals differ only in their skill level (denoted by n) which measures their marginal productivity. If an individual with skill n supplies labor or “effort” l , he earns nl . The distribution of skills is written $F(n)$, with density $f(n)$ and support in $[0, \infty)$. f is also assumed to be regular (at least of class C^2). The consumption choice of an individual with skill n is denoted by (c_n, l_n) and I write $z_n = nl_n$ for its earnings and u_n for its utility level $u(c_n, l_n)$. The government does not observe n or l_n but only earnings z_n . Thus it is restricted to setting taxes as a function only of earnings: $c = z - T(z)$. The government maximizes the following social welfare function:

$$W = \int_0^{\infty} G(u_n)f(n)dn \quad (1.16)$$

where G is an increasing and concave function of utility. The government maximizes W subject to a resource constraint and an incentive compatibility constraint. The resource constraint states that aggregate consumption must be less than aggregate production minus government expenditures, E :

$$\int_0^{\infty} c_n f(n)dn \leq \int_0^{\infty} z_n f(n)dn - E \quad (1.17)$$

The incentive compatibility constraint is that the selected labor supply l_n maximizes utility, given the tax function, $u(nl - T(nl), l)$. Assuming that the tax schedule T is regular, the optimal choice of l implies the following first order condition:

$$n(1 - T'(z_n))u_c + u_l = 0 \quad (1.18)$$

This equation holds true as long as the individual chooses to supply a positive amount of labor. This first order condition leads to:

$$\dot{u}_n = -\frac{l u_l}{n} \quad (1.19)$$

where a dot means (total) differentiation with respect to the skill level n .

Following Mirrlees (1971), in the maximization program of the government, u_n is regarded as the state variable, l_n as the control variable while c_n is determined implicitly as a function of u_n and l_n from the equation $u_n = u(c_n, l_n)$. Therefore, the program of the government is simply to maximize equation (1.16) by choosing l_n and u_n subject to equations (1.17) and (1.19). Forming a Hamiltonian for this expression, we have:

$$H = [G(u_n) - p(c_n - nl_n)]f(n) - \phi(n)\frac{l_n u_l(c_n, l_n)}{n} \quad (1.20)$$

where p and $\phi(n)$ are multipliers. p is the Lagrange multiplier of the government's budget constraint and thus can be interpreted as the marginal value of public funds. From the first order conditions of maximization, we obtain the classical first order condition for optimal rates (see Mirrlees (1971), equation (33)):

$$(n + \frac{u_l^{(n)}}{u_c^{(n)}})f(n) = \frac{\psi_l^{(n)}}{n} \int_n^\infty [\frac{1}{u_c^{(m)}} - \frac{G'(u_m)}{p}] T_{nm} f(m) dm \quad (1.21)$$

where $T_{nm} = \exp[-\int_n^m \frac{t_s u_{cl}(c_s, l_s)}{s u_c(c_s, l_s)} ds]$. ψ is defined such that $\psi(u, l) = -lu_l(c, l)$ where c is a function of (u, l) such that $u = u(c, l)$. An superscript (n) means that the corresponding function is estimated at (c_n, l_n, u_n) . The derivation of (1.21) is recalled in appendix.

Theorem 2 in Mirrlees (1971) (pp.183-4) states under what conditions formula (1.21) is satisfied at the optimum. The most important assumption is the single-crossing condition (condition (B) in Mirrlees (1971)) which is equivalent to the uncompensated elasticity being greater than minus one. This condition is very likely to hold empirically and I will also assume from now on that $\zeta^u > -1$.

Even when the single crossing property is satisfied, the first order condition (1.21) may not characterize the optimum. The complication comes from the need to check that individual labor supply choices satisfying the first order condition (1.18) are globally optimal choices. Mirrlees showed that the first order condition for individual maximization implies global maximization if and only if the earnings function z_n is non-decreasing in the skill level n . If equation (1.21) leads to earnings z_n decreasing over some skill ranges then this cannot be the optimum solution and therefore there must be bunching at some income level (a range of workers with skills n lying in $[n_1, n_2]$ choosing the same income level \bar{z}). When bunching happens, (1.21) no longer holds but $\dot{u}_n = -lu_l/n$ remains true. Theorem 2 in Mirrlees (1971) states that (1.21) holds at every point n where z_n is increasing.¹⁸ Seade (1982) showed that if leisure is not an inferior good (i.e., $mpe \leq 0$) then T' cannot be negative at the optimum. I assumed in Section 3 that $mpe \leq 0$ and continue to do so in this Section.

In this model, redistribution takes place through a guaranteed income level that is taxed away as earnings increase (negative income tax). Optimal marginal tax rates are defined by equation (1.21). Therefore, the welfare program is fully integrated to the tax program.

¹⁸Gaps in the distribution of incomes can also happen in case of multiple maxima in the maximization of the Hamiltonian with respect to l . Gaps do not arise generically and can be ruled out under weak assumptions (see Mirrlees (1971)). I will therefore assume from now on, that the equilibrium distribution of incomes has no gaps.

1.4.2 Optimal Marginal Rates

The general Mirrlees first order condition (1.21) depends in a complicated way on the derivatives of the utility function $u(c, l)$ which are almost impossible to measure empirically. Therefore, it has been impossible to infer directly from the general equation (1.21) practical quantitative results about marginal rate patterns. Moreover, equation (1.21) has always been derived using powerful but blind Hamiltonian optimization. Thus, the optimal taxation literature has never been able to elucidate the key economic effects which lead to the general formula (1.21). In this subsection, I rewrite equation (1.21) as a function of elasticities of earnings and show precisely the key behavioral effects which lead to this rewritten equation. I first present a simple preliminary result that is a useful step to understand the relation between the income distribution and the distribution of skills in the Mirrlees economy.

Lemma 1 *For any regular tax schedule T (such that T'' exists) not necessarily optimal, the earnings function z_n is non-decreasing and satisfies the following equation,*

$$\frac{\dot{z}_n}{z_n} = \frac{n\dot{l}_n + l_n}{nl_n} = \frac{1 + \zeta_n^u}{n} - \dot{z}_n \frac{T''(z_n)}{1 - T'(z_n)} \zeta_n^c \quad (1.22)$$

If equation (1.22) leads to $\dot{z}_n < 0$ then z_n is discontinuous and (1.22) does not hold.

The proof, which is routine algebra, is presented in appendix. In the case of a linear tax ($T'' = 0$) the earnings equation (1.22) becomes the familiar equation $dz/z = (1 + \zeta^u)dn/n$. In the general case, a correction term in T'' which represents the effect of the change in marginal rates is present.

The first order condition (1.21) can be reorganized in order to express optimal tax rates in terms of the elasticities of earnings. This rearrangement of terms is a generalization of the one introduced in Diamond (1998) in the case of quasi-linear utility functions.

Proposition 1 *The first order condition (1.21) can be rewritten as follows:*

$$\frac{T'(z_n)}{1 - T'(z_n)} = A(n)B(n) \quad (1.23)$$

where

$$A(n) = \left(\frac{\zeta_{(n)}^u + 1}{\zeta_{(n)}^c} \right) \left(\frac{1 - F(n)}{nf(n)} \right) \quad (1.24)$$

$$B(n) = \int_n^\infty \left[1 - \frac{G'(u_m)u_c^{(m)}}{p} \right] S_{nm} \frac{f(m)}{1 - F(n)} dm \quad (1.25)$$

where

$$S_{nm} = \exp \left[\int_n^m \left(1 - \frac{\zeta_{(s)}^u}{\zeta_{(s)}^c} \right) \frac{dz_s}{z_s} \right] \quad (1.26)$$

The formal proof of this proposition, which starts with equation (1.21) and is routine algebra, is presented in appendix. This proof, however, does not show the economic effects which lead to formula (1.23). It is possible, though, to derive this equation by considering small variations in marginal rates around the optimum as in Section 3. This derivation, though complicated, shows precisely how the key effects come into play to lead to formula (1.23) and therefore is presented in detail. Formula (1.23) is commented in the light of this direct derivation just after the proof.

Direct Proof of Proposition 1

I note $H(z)$ the distribution of incomes at the optimum and $h(z)$ the corresponding density function. I note again $MS(z)$ the marginal social value of consumption for a taxpayer with income z (i.e., this is exactly $G'(u)u_c$ is Mirrlees notation). p is the marginal value of public funds. I consider the effect of the following small tax reform: marginal rates are increased by an amount $d\tau$ for incomes between \bar{z} and $\bar{z} + d\bar{z}$.¹⁹

This tax reform has three effects on tax receipts: a mechanical effect, an elasticity effect for taxpayers with income between \bar{z} and $\bar{z} + d\bar{z}$, and an income effect for taxpayers with income above \bar{z} .

• Mechanical Effect

This effect represents the increase in tax receipts if there were no behavioral responses. Every taxpayer with income z above \bar{z} pays $d\tau d\bar{z}$ additional taxes which are valued $(1 - MS(z)/p)d\tau d\bar{z}$ by the government therefore the overall net effect M

¹⁹I also assume that $d\tau$ is second order compared to $d\bar{z}$ so that bunching (and inversely gaps in the income distribution) around \bar{z} or $\bar{z} + d\bar{z}$ induced by the discontinuous change in marginal rates are negligible.

is equal to:²⁰

$$M = d\tau d\bar{z} \int_{\bar{z}}^{\infty} \left[1 - \frac{MS(z)}{p}\right] h(z) dz$$

• **Elastic Effect**

The increase $d\tau$ for a taxpayer with income z between \bar{z} and $\bar{z} + d\bar{z}$ has an elastic effect which produces a small change in income (denoted by dz). This change is the consequence of two effects. First, there is a direct compensated effect due to the exogenous increase $d\tau$. The compensated elasticity is the relevant one here because the change $d\tau$ takes place at level \bar{z} just below z (see the discussion following equation (1.7) in the previous Section). Second, there is an indirect effect due to the shift of the taxpayer on the tax schedule by dz which induces an endogenous additional change in marginal rates equal to $dT' = T'' dz$. Therefore, the behavioral equation can be written as follows,

$$dz = -\zeta^c \bar{z} \frac{d\tau + dT'}{1 - T'}$$

which implies,

$$dz = -\zeta^c \bar{z} \frac{d\tau}{1 - T' + \zeta^c \bar{z} T''}$$

By Lemma 1, $1 - T' + \zeta^c \bar{z} T'' = (1 - T')(1 + \zeta^u)z / (n\dot{z}) > 0$. When the Single Crossing condition $1 + \zeta^u > 0$ holds, $1 - T' + \zeta^c \bar{z} T'' > 0$ if and only if $\dot{z} > 0$. As reviewed above, $\dot{z} \geq 0$ is a necessary and sufficient condition for the individual choice given by the individual first order condition to be a global maximum. Thus $1 - T' + \zeta^c \bar{z} T'' \geq 0$ is also necessary and sufficient to insure global optimization of the individual choice. I assume in this heuristic proof that $1 - T' + \zeta^c \bar{z} T'' > 0$ for any \bar{z} in order to avoid dealing with bunching issues.²¹

In order to simplify notations, I introduce $h^*(\bar{z})$ which is the density of incomes that would take place at \bar{z} if the tax schedule $T(\cdot)$ were replaced by the linear tax

²⁰The tax reform has also an effect on $h(z)$ but this is a second order effect in the computation of M .

²¹This condition is always satisfied at points where $T''(\bar{z}) \geq 0$.

schedule tangent to $T(\cdot)$ at level \bar{z} .²² I call the density $h^*(\bar{z})$ the virtual density. Densities h and h^* are related through the skill density $f(n)$ such that $h^*(\bar{z})\dot{z}^* = h(\bar{z})\dot{z} = f(n)$ where \dot{z}^* is the derivative of earnings with respect to n at point \bar{z} if the tax schedule T is replaced by the tangent linear tax schedule. Using Lemma 1, I have, $\dot{z}^*/\bar{z} = (1 + \zeta^u)/n$ and $\dot{z}/\bar{z} = (1 + \zeta^u)/n - \dot{z}\zeta^c T''/(1 - T')$ which implies:

$$\frac{h^*(\bar{z})}{1 - T'(\bar{z})} = \frac{h(\bar{z})}{1 - T'(\bar{z}) + \zeta_{(\bar{z})}^c \bar{z} T''(\bar{z})} \quad (1.27)$$

where $\zeta_{(\bar{z})}^c$ is the compensated elasticity at income level \bar{z} . Using $h^*(\bar{z})$, the overall effect on tax receipts (denoted by E) can be simply written as:

$$E = -\zeta_{(\bar{z})}^c \bar{z} \frac{T'}{1 - T'} h^*(\bar{z}) d\tau d\bar{z}$$

• Income Effect

A taxpayer with income z above \bar{z} pays $-dR = d\tau d\bar{z}$ additional taxes. This produces an income response dz which is again due to two effects. First, there is the direct income effect (equal to $mpe dR/(1 - T')$). Second, there is an indirect elastic effect due to the change in marginal rates $dT' = T'' dz$ induced by the shift dz along the tax schedule. Therefore,

$$dz = -\zeta^c z \frac{T'' dz}{1 - T'} - mpe \frac{d\tau d\bar{z}}{1 - T'}$$

which implies,

$$dz = -mpe \frac{d\tau d\bar{z}}{1 - T' + z\zeta^c T''} \quad (1.28)$$

Introducing again the density $h^*(z)$ and summing (1.28) over all taxpayers with income larger than \bar{z} , I obtain the total behavioral effect I due to income effects:

$$I = d\tau d\bar{z} \int_{\bar{z}}^{\infty} -mpe_{(z)} \frac{T'}{1 - T'} h^*(z) dz$$

At the optimum, the sum of the three effects M , E and I must be zero which implies,

²²This linear tax schedule is characterized by the tax rate $\tau = T'(\bar{z})$ and the virtual income $R = \bar{z} - T(\bar{z}) - \bar{z}(1 - \tau)$.

$$\frac{T'}{1-T'} = \frac{1}{\zeta^c} \left(\frac{1-H(\bar{z})}{\bar{z}h^*(\bar{z})} \right) \left[\int_{\bar{z}}^{\infty} \left(1 - \frac{MS(z)}{p} \right) \frac{h(z)}{1-H(\bar{z})} dz + \int_{\bar{z}}^{\infty} -mpe \frac{T'}{1-T'} \frac{h^*(z)}{1-H(\bar{z})} dz \right] \quad (1.29)$$

Equation (1.29) can be considered as a first order linear differential equation and can be integrated (see appendix) using the standard method to obtain:

$$\frac{T'(\bar{z})}{1-T'(\bar{z})} = \frac{1}{\zeta_{(\bar{z})}^c} \left(\frac{1-H(\bar{z})}{\bar{z}h^*(\bar{z})} \right) \int_{\bar{z}}^{\infty} \left(1 - \frac{MS(z)}{p} \right) \exp \left[\int_{\bar{z}}^z \left(1 - \frac{\zeta_{(z')^u}}{\zeta_{(z')^c}} \right) \frac{dz'}{z'} \right] \frac{h(z)}{1-H(\bar{z})} dz \quad (1.30)$$

Changing variables from \bar{z} to n , and using the fact, proved above, that $\bar{z}h^*(\bar{z})(1+\zeta^u) = nf(n)$, it is straightforward to obtain the equation of Proposition 1. Therefore, when changing variables from \bar{z} to n , an additional term $1 + \zeta^u$ appears on the righthand side to form the term $A(n)$ of Proposition 1. This counterintuitive term (higher uncompensated elasticity should not lead to higher marginal rates) should in fact be incorporated into the skill ratio $(1-F)/(nf)$ to lead to the income ratio $(1-H)/(\bar{z}h^*)$ which is easier to relate to the empirical income distribution. Of course, the virtual density h^* is not identical to the actual density h . However, because the density h at the optimum is endogenous (in the sense that changes in the tax schedule affect the income distribution), there is very little inconvenience in using h^* rather than h . Using Lemma 1, one can observe that nonlinear tax schedules produce a deformation of the earnings distribution h . Using h^* is a way to get rid of this deformation component. In that sense, h^* is more closely related than h to the skill distribution which represents intrinsic inequalities.

Last, let me mention that the multiplier p is such that the integral term in (1.30) must be zero when $\bar{z} = 0$. This can be proved by considering that a small change in the lump sum given to everybody ($-T(0)$) has no first order effect on total welfare.

QED

Interpretation of Proposition 1

In the light of this heuristic proof, let us analyze the decomposition of optimal tax rates presented in Proposition 1 or equivalently equation (1.30). Analyzing equation

(1.30) (or (1.23)), it appears that three elements determine optimal income tax rates: elasticity (and income) effects, the shape of the income (or skill) distribution and social marginal weights. These elements enter the optimal tax formula in relatively independent ways and thus can be examined independently.

• **Shape of Income Distribution**

The shape of the income distribution affects the optimal rate at level \bar{z} mainly through the term $(1 - H(\bar{z})) / (\bar{z}h(\bar{z}))$. This is intuitive: the elastic distortion at \bar{z} induced by a marginal rate increase at that level is proportional to income at that level times number of people at that income level ($\bar{z}h(\bar{z})$) while the gain in tax receipts is proportional to the number of people above \bar{z} (i.e., $1 - H(\bar{z})$). In other words, a high marginal rate at a given income level \bar{z} induces a negative behavioral response at that level but allows the government to raise more taxes from all taxpayers above \bar{z} . Therefore, the government should apply high marginal rates at levels where the density of taxpayers is low compared to the number of taxpayers with higher income. Unsurprisingly, the ratio $(1 - H)/(zh)$ is constant and equal to $1/a$ when $H(z)$ is Pareto distributed with parameter a . This ratio tends to zero when the top tail is thinner than any Pareto distribution. Next subsection examines the asymptotics of formula (1.30). The empirical shape of the ratio $(1 - H)/(zh)$ is studied in Section 5.

• **Elastic and Income effects**

Behavioral effects enter the formula for optimal rates in two ways. First, increasing marginal rates at level \bar{z} induces a compensated response from taxpayers earning \bar{z} . Therefore, $\zeta_{(\bar{z})}^c$ enters negatively the optimal tax rate at income level \bar{z} . Second, this marginal rate change increases the tax burden of all taxpayers with income above \bar{z} . This effect induces these taxpayers to work more through income effects which is good for tax receipts. Therefore, this income effect leads to higher marginal rates (everything else being equal) through the term S_{nm} (or equivalently the exponential term in (1.30)) which is bigger than one.²³ Note that this term is identically equal to one when there are no income effects (this case was studied by Diamond (1998)).

²³The term $1 - G'(u)u_c/p = 1 - MS(z)/p$ is in general increasing in income and is thus always positive above some income level.

The heuristic proof shows clearly why negative tax rates are never optimal. If the tax rate were negative in some range then increasing it a little bit in that range would decrease earnings in that range (because of the substitution effect) but this behavioral response would increase tax receipts because the tax rate is negative in that range. Therefore, this small tax reform would unambiguously increase welfare.

• Social Marginal Welfare Weights

The social marginal weights (denoted by $MS(z)/p$ in terms of the marginal value of public funds) enter the optimal tax formula through the term $(1 - MS(z)/p)$ inside the integral. The intuition is the following: increasing marginal rates locally at level \bar{z} increases the tax burden of all taxpayers with income above \bar{z} . Each additional dollar raised by the government over taxpayers with income z is valued $(1 - MS(z)/p)$. This expression is decreasing with z (as long as the government has redistributive goals). Therefore, redistributive goals is unsurprisingly an element tending to make the tax schedule progressive. If the government had no redistributive goals, then it would choose the same marginal welfare weights for everybody. The formula for the optimal income tax would clearly be qualitatively very close to the general case with redistributive concerns. In particular, the shape of the income distribution and the size of both substitution and income effects would matter for the optimal income tax with no redistributive goals.²⁴

The original Mirrlees's derivation relies heavily on the fact that there exists a unidimensional skill parameter which characterizes each taxpayer. As a result, that derivation gives no clue about how to extend the non-linear tax formula to a heterogeneous population in a simple way. The direct proof using elasticities shows that there is no need to introduce an exogenous skill distribution. Formula (1.30) is valid for any heterogeneous population as long as $\zeta_{(z)}^u$ and $\zeta_{(z)}^c$ are considered as average elasticities at income level z .²⁵ Therefore, the skill distribution in the Mirrlees model should not be considered as a real economic element (which one should try to measure

²⁴Chapter 2 investigates this point more deeply. I show that the income tax which minimizes deadweight burden is in fact an optimal income tax with particular welfare weights. In the absence of income effects, these weights are the same for everybody.

²⁵Equation (1.27) linking the virtual density h^* to the actual density h can be generalized to the case of heterogeneous populations.

empirically) but rather as a useful simplification device to perform computations and numerical simulations. The skill distribution should simply be chosen so that the resulting income distribution be close to the empirical income distribution.²⁶ Mirrlees (1976) and (1986) tried to extend his 1971 formula to heterogeneous populations where individuals are characterized by a multidimensional parameter instead of a single dimensional skill parameter. He adopted the same approach as he used in his 1971 study and derived first order conditions for the optimal tax schedule. However, these conditions were even more complicated than in the unidimensional case and thus it proved impossible to obtain results or interpret the first order conditions in that general case. It is nonetheless possible to manipulate the first order conditions of the general case considered in Mirrlees (1976) and (1986) in order to recover formula (1.30). Therefore, the elasticity method of the heuristic proof is a powerful tool to understand the economics of optimal income taxation and is certainly a necessary step to take to extend in a fruitful way the model to heterogeneous populations. This general derivation is out of the scope of the present paper and will be presented in future work.

Though this is not attempted in this paper, let me sketch how formula (1.30) could be used to perform numerical simulations without need to rely on an exogenous skill distribution. Making assumptions about the pattern of elasticities,²⁷ selecting a function $MS(\cdot)$ reflecting the redistributive tastes of the government, and using the empirical income distribution to obtain $H(\cdot)$, equation (1.30) could be used to compute a tax schedule $T'(\cdot)$. Of course, this tax schedule would not be optimal because $H(\cdot)$ is an endogenous function (a tax reform affects income distribution through behavioral responses). Nevertheless, this computed $T'(\cdot)$ could yield interesting information for tax reform. Using this estimated T' , a new income distribution $H(\cdot)$ could then be derived leading to a new estimate for T' . This algorithm may converge to the optimal tax schedule. This avenue of research is out of the scope of the present paper but may deserve further investigation.

Formula (1.30) could also be used to pursue a *positive* analysis of actual tax schedules. Considering the actual tax schedule $T(\cdot)$ and the actual income distribution

²⁶This route is followed in Section 5.

²⁷I review in Section 5 the empirical results about these elasticities.

$H(\cdot)$, and making assumptions about the patterns of elasticities $\zeta_{(z)}^u$ and $\zeta_{(z)}^c$, it is also possible to use equation (1.30) to infer the marginal social weights $MS(z)/p$. Even if the government does not really 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. Alternatively, a government maximizing median voter utility would choose for the weights $MS(\cdot)$ a Dirac distribution centered at the median income level, this would produce a jump in marginal tax rates at the median income level. This type of analysis could also be used to assess how different tax reform proposals would map into a change in the weights $MS(z)/p$. This line of research is left for future work.

The remaining part of this Section examines the asymptotics of optimal marginal rates in the framework of the Mirrlees model. I first examine the link between the skill distribution and the income distribution (understanding this link is crucial to perform the numerical simulations of Section 5 using the empirical earnings distribution). I then rederive the formula for high income optimal rates of Section 3 by examining the asymptotics of the general formula for optimal rates discussed above. Readers less interested in technicalities can skip Section 4.3 and go directly to Section 5.

1.4.3 Optimal Asymptotic Rates

From the skill distribution to the income distribution

Section 3 has shown that Pareto distributions provide a benchmark of central importance to understand optimal asymptotic rates. The optimal rate depends on the limiting behavior of the tail of the income distribution. This limiting behavior can be characterized by a limiting “Pareto” parameter. This subsection first defines this limiting “Pareto” behavior in a rigorous way. I then show that the limiting “Pareto” parameters of the skill distribution and of the earnings distribution are linked through the asymptotic uncompensated elasticity.

$F(n)$ is called a Pareto distribution with parameter γ if and only if $F(n) = 1 - C/n^\gamma$ for some constant C . Its density function is equal to $f(n) = \gamma C/n^{1+\gamma}$. A Pareto density is always decreasing while empirical distributions are in general unimodal (first increasing and then decreasing). Therefore, Pareto distributions are

useful to approximate empirical distributions only above the mode. If $F(n)$ is a Pareto distribution with parameter γ , then:

$$nf'(n)/f(n) = -(1 + \gamma) \quad (1.31)$$

For any $\alpha < \gamma$,

$$\int_n^\infty m^\alpha f(m)dm/[n^\alpha(1 - F(n))] = \gamma/(\gamma - \alpha) \quad (1.32)$$

In particular, the mean above any level \bar{n} divided by \bar{n} (i.e., the conditional mean ratio $E(n|n > \bar{n})/\bar{n}$) is constant and equal to $\gamma/(\gamma - 1)$. From these two properties characterizing Pareto distributions, I consider two corresponding definitions of the asymptotic relationship between any given distribution and a Pareto distribution.

If $F(n)$ is a regular (at least C^2) distribution function with support in $[0, +\infty)$ and density function $F'(n) = f(n)$ such that (for some $\gamma > 0$ possibly infinite):

$$\lim_{n \rightarrow \infty} nf'(n)/f(n) = -(1 + \gamma) \quad (1.33)$$

then I say that F behaves *strongly* like a Pareto distribution with parameter γ .

If $F(n)$ is a distribution that satisfies:

$$\lim_{n \rightarrow \infty} \int_n^\infty m^\alpha f(m)dm/[n^\alpha(1 - F(n))] = \gamma/(\gamma - \alpha) \quad (1.34)$$

for any $\alpha < \gamma$ then I say that F behaves *weakly* like a Pareto distribution with parameter γ .

These definitions are constructed in such a way that property (1.33) implies (1.34). The reverse implication is not necessarily true if f is not regular enough. The proof is easy and presented in appendix. Now, the following proposition linking the skill distribution and the income distribution can be proved.

Proposition 2 *Suppose that the distribution of skills $f(n)$ behaves weakly like a Pareto distribution (property (1.34)) with parameter γ (possibly infinite). Suppose that the tax rate schedule (not necessarily optimal) T' tends to $\bar{\tau} < 1$ as n tends to*

infinity. Suppose also that T' is such that there is no bunching nor gaps above some income level.

Suppose that the compensated and uncompensated elasticities converge to $\bar{\zeta}^c \geq 0$ and $\bar{\zeta}^u > -1$ as n tends to infinity. In the case $\bar{\zeta}^c = 0$ assume in addition that $\zeta_{(n)}^c \downarrow 0$ for n large.

Then the distribution of earnings behaves weakly like a Pareto distribution (property (1.34)) with parameter $a = \gamma/(1 + \bar{\zeta}^u)$.

The formal proof is presented in appendix. The idea of the proof is easy to understand. From the definition of the uncompensated elasticity, we have $dl/l = \zeta^u dw/w$. Assuming that taxpayers face a linear tax schedule (with constant virtual income R and net of tax wage rates $w = n(1 - \tau)$) and that ζ^u is constant and equal to $\bar{\zeta}^u$, it is possible to integrate the above equation over wage rates to obtain $l(n) \simeq Cn^{\bar{\zeta}^u}$ which implies $z_n = nl(n) \simeq Cn^{1+\bar{\zeta}^u}$. If the wage rates are roughly Pareto distributed above a given wage rate level (i.e., $Prob(n > \bar{n}) \simeq C/\bar{n}^\gamma$), then:

$$Prob(\text{income} > z) \simeq Prob(Cn^{1+\bar{\zeta}^u} > z) = Prob(n > (z/C)^{\frac{1}{1+\bar{\zeta}^u}}) \simeq C'/z^a$$

where $a = \gamma/(1 + \bar{\zeta}^u)$. Therefore the distribution of incomes is also roughly Pareto distributed with parameter $a = \gamma/(1 + \bar{\zeta}^u)$ instead of γ .

This result is important because while it is very difficult to observe distributions of skills, observing empirical distributions of wages is much easier. It must be noted that the ‘‘Pareto’’ parameter of the income distribution does not depend on the limiting tax rate $\bar{\tau}$. Therefore, a can be inferred directly from the observation of empirical earnings distributions. Surprisingly, the optimal taxation literature has not noticed this simple result. This may explain why researchers did not try to calibrate numerical simulations to empirical income distributions. They almost always used log-normal skill distributions which match roughly unimodal empirical distributions but approximate very poorly empirical distributions at the tails (both top and bottom tails). Moreover, changing the elasticity parameter without changing the skill distribution, as usually done in numerical simulations, might be misleading. As evidenced in Proposition 2, changing the elasticities modifies the resulting income distribution

and thus might affect optimal rates also through this indirect effect. I come back to this point in Section 5.

Asymptotic Rates

Proposition 3 *Assume that along the optimal tax schedule, the elasticities $\zeta_{(n)}^c$ and $\zeta_{(n)}^u$ and income effects $mpe_{(n)}$ converge to values denoted by $\bar{\zeta}^c$, $\bar{\zeta}^u$ and \overline{mpe} when n tends to infinity. Assume that $\overline{mpe} > -1$ (and therefore $\bar{\zeta}^u > -1$). Assume that the ratio of the social marginal utility to marginal value of public funds $G'(u_n)u_c^{(n)}/p$ converges to \bar{g} as n tends to infinity. Assume that the distribution of skills $f(n)$ behaves strongly like a Pareto distribution (property (1.33)) with parameter γ (possibly infinite). Assume also that there is no bunching above some income level.*

Then, the optimal tax rate T' tends to a limit $\bar{\tau}$ such that:

$$\bar{\tau} = \frac{1 - \bar{g}}{1 - \bar{g} + \bar{\zeta}^u + \bar{\zeta}^c(a - 1)} \quad (1.35)$$

where, $a = \gamma/(1 + \bar{\zeta}^u)$ is the ‘‘Pareto’’ parameter of the tail of the income distribution (as in Proposition 2). If formula (1.35) leads a value bigger than one for $\bar{\tau}$ then it must be understood that T' tends to one.

The full proof is in appendix. However, using Proposition 1 and 2 and Lemma 1, I can give an idea of the proof. If we admit that T' converges then the term involving T'' in (1.22) becomes negligible and therefore dz_s/z_s can be replaced by $(1 + \zeta^u)ds/s$ in (1.26). Now assuming that $G'(u)u_c/p$ is constant and equal to \bar{g} , that the elasticities are constant, and that $f(n)$ is exactly Pareto distributed, straightforward calculations using (1.31) and (1.32) show that T' is exactly equal to $\bar{\tau}$ of Proposition 3.

Therefore, this sketch shows that if the skills are exactly Pareto distributed, the elasticities exactly constant, and the social marginal value constant above a given level of skills then the government would apply a constant marginal rate above this level of skills. Thus formula (1.35) is likely to be relevant over a broad range of incomes. I come back to this issue in more detail in Section 5.

1.5 Empirical Results and Simulations

This Section is divided in two parts. First, I examine empirical distributions of wages and discuss elasticity estimates found in the applied literature in order to present asymptotic optimal rates for a range of realistic parameters. Second, I perform numerical simulations to compute optimal tax schedules in the Mirrlees model using empirical earnings distributions.

1.5.1 Optimal high income tax rates

Empirical elasticities

Labor supply studies have consistently found small or negative uncompensated elasticities of male hours of work (see Pencavel (1986), p.69 and p.73). These studies find in general uncompensated elasticities slightly below 0 (around -0.1) and compensated elasticities slightly higher than zero (around 0.1). Non-linear budget set studies which tend to find larger compensated elasticities have also found small uncompensated elasticities (see Hausman (1985), p.241). The estimates for uncompensated elasticities are also around 0 but the compensated elasticities are usually between 0.2 and 0.5. The labor supply elasticity of women has been found in general higher than the one for men (e.g. Eissa (1995)). Elasticity estimates range in general from 0.5 to 1. However, it should be noticed that the relevant elasticity for a couple is the average elasticity with weights equal to the income share of each member. Even if the elasticity of the second earner is high, the total elasticity of the couple is likely to remain small because the share of the second earner's income is usually small.

Nevertheless, we have seen that for the optimal tax problem, what matters is the total elasticity of earnings and not only the elasticity of hours of work. The former should be higher than the later because hours of work are not the only dimension of "effort". Individuals can vary their labor supply not only by changing hours but also the intensity of work or the types of job they enter in²⁸. Several recent empirical studies have found large elasticities of taxable income with respect to net of tax rates (Lindsey (1987), Feldstein (1995), Navratil (1995) and Auten and Carroll (1997)).

²⁸Feldstein (1995) explains this point in more detail.

The elasticities estimated by these authors are around (or even above) one.

These high elasticity results have been criticized on several grounds. First, these studies compare the increase in incomes of high income earners (who experienced large marginal rate cuts) to the increase in incomes of middle or low income earners (who experienced much smaller marginal tax cuts). This methodology amounts therefore to attributing the widening in inequalities to the tax reform. Second, the tax cuts of the 1980s introduced many changes in tax rules which affected the incentives for reporting taxable income. In particular, the incentives for shifting labor income to capital income or for shifting personal income to corporate income may have been substantially reduced by the tax reforms. This issue is investigated in Auerbach and Slemrod (1997) and Gordon and Slemrod (1999). Chapter 3 estimates compensated elasticities of reported income with respect to tax rates using the bracket creep in the US from 1979 to 1981. Although this tax change induced smaller tax rate changes than the tax reforms of the 1980s, it does not suffer from the two problems mentioned. This study finds much lower income elasticity estimates between 0 and 0.5. Last, the tax cuts studies are unable to distinguish between permanent shifts to the form of compensation and temporary shifts to the timing of compensation. This issue was pointed out in Slemrod (1995). Goolsbee (1997) investigates this point using the tax rate increases for high income earners enacted in 1993 and compensation data on corporate executives from 1991 to 1995. He shows convincingly that the tax reform led to a large income shifting from 1993 to 1992 to escape higher tax rates, implying a very large short term elasticity (above one); however, the elasticity after one year is small (at most 0.4 and probably close to zero).

Contrary to most labor supply studies, tax reform studies are in general unable to estimate both substitution and income effects. The elasticities estimated are therefore a mix of compensated and uncompensated effects. In summary, the elasticities of total earnings for high income earners are still poorly known. They are likely to be smaller than those found in the studies of Lindsey (1987) and Feldstein (1995) and may not be significantly larger than those of middle income earners.

Empirical wage income distributions

Section 4.1 showed that the conditional mean income ratio (i.e. $E(z|z > \bar{z})/\bar{z}$) is an important element for optimal tax rates. I have computed this function using data on wage earnings from individual tax returns. The Internal Revenue Service (IRS) constructs each year a large cross-section of tax returns (about 100,000 observations per year). These datasets overweight wealthy taxpayers and therefore are one of the most valuable source of information about high income earners. As almost all wealthy taxpayers are married filing jointly, I focus only on this class of taxpayers. As I consider taxation of labor income, I focus mostly on wage income.²⁹ I define narrowly wage income as income reported on the line “wages, salaries and tips” of the US income tax form.

Figures 2 and 3 plot the values of the conditional mean income ratios as a function of \bar{z} for two different ranges of income. Figure 2 is for incomes between 0 to 500,000 dollars (all Figures are expressed in 1992 dollars and represent yearly income) and Figure 3 for incomes between \$10,000 to \$30 million using a semi-log scale. The Figures show that the conditional mean income ratio is strikingly stable over the tail of the income distribution. The value is around 2.3 for 1992 and 2.1 for 1993. If anything, the curve seems to be slightly increasing from \$100,000 to \$5 million. The plots on Figure 3 become noisy above \$10 million because the number of taxpayers above that level is very small and crossing only one taxpayer has a non trivial discrete effect on the curves. As discussed in Section 3, the ratio must be equal to one at the level of the highest income. However, Figure 3 shows that even at income level \$30 million, the ratio is still around 2. For example, if the second top income taxpayer earns half as much as the top taxpayer then the ratio is equal to 2 at the level of the second top earner. Consequently, the zero top result only applies to the very highest taxpayer and is therefore of no practical interest. Empirical distributions give much support to the assumption that the conditional mean income ratio converges as income increases. In fact, above 150,000 dollars, this ratio can be considered as

²⁹It is well known that wealthy taxpayers tend to shift labor income to capital income in order to pay less taxes (see Slemrod (1996)). Note however that after the Tax Reform Act of 1986 and until the tax increases of 1993, tax rates on labor and capital were very similar and therefore the incentives for income shifting were probably much lower than they had been before.

roughly constant and thus the theory developed in Section 3 is relevant over a broad range of incomes. As seen in Section 3, nearly constant conditional mean income ratio means that the income distribution can be well approximated by a Pareto distribution with parameter $a = (z_m/\bar{z})/(z_m/\bar{z} - 1)$ and therefore formula (1.12) can be applied. Pareto parameters for the wage income distribution are estimated between 1.8 and 2.2 (depending on years³⁰).

The mean ratio declines quickly until \$60,000 and then increases from 1.7 to 2.2 until \$130,000. Therefore, if elasticities were roughly constant above \$60,000, the results of Section 3 show that the optimal linear tax rate $\bar{\tau}$ that the government would like to set above the income level \bar{z} is increasing over the range 60,000 to 130,000 dollars. This suggests that the optimal non-linear tax rate is likely to be increasing over that range. I examine this point in detail later on.

The IRS has constructed tax returns files since year 1960. Therefore, it is possible to plot the conditional mean income ratios for many different years and various types of incomes. Because of limited space, I present only two additional Figures. On Figure 4, I plot conditional mean income ratios for years 1987 to 1993 and wage income between 0 and 1,000,000 dollars (incomes are expressed in 1992 dollars). The vertical scale has been expanded so as to stress the differences between the different years. The conditional mean income ratios vary from year to year from a low 1.85 (in 1987) to a high 2.25 (in 1992). In year 1987, the TRA of 1986 was not yet fully phased in and the top tax rate was 38.5% (instead of 28% in 1988). From 1988 to 1992, the top rate was relatively stable (28% in 1988 and 1989 and 31% in 1990, 1991 and 1992). In 1993, the top rate was increased to 39.6% (the top rate for capital gains remained at 28%). The ratio is the lowest for 1987 and one of the highest for 1988, suggesting income shifting from 1987 to 1988 to avoid the high 1987 top rate. The ratio for 1992 (which was the last year before OBRA 1993 significantly increased the top rate) is the highest one, suggesting again a shift from 1993 to 1992 to avoid high rates.

Figure 5 presents the same plots for Adjusted Gross Income (AGI is a measure of total income including both capital and labor income). The ratios are higher than for

³⁰Feenberg and Poterba (1993) have estimated Pareto parameters between 1.5 and 2.5 for the top distribution of Adjusted Gross Income over the period 1951-1990.

wages (from a low 2.4 to a high 2.7). 1987 is one of the lowest years.³¹ Year 1988 is by far the highest year, supporting the shifting interpretation. The difference between 1992 and 1993 is much smaller for AGI than for wages. The 1993 tax increase did not affect capital gain taxes and thus shifting labor income toward capital income may have decreased the conditional mean income ratio for wages without much affecting the AGI ratio. Looking at the conditional mean income ratios provides interesting information about high income taxpayers' responses to marginal rates and suggests that most of the response is due to short run intertemporal shifts of income around tax reforms years. Extending this study to other years and other types of incomes is left for future research.

It is also interesting to plot the empirical ratio $[1 - H(z)]/zh(z)$, which I call from now on the hazard ratio. This ratio has been shown to be highly relevant for computing optimal tax rates in the general non-linear case. This ratio is exactly equal to $1/a$ if $H(z)$ is Pareto distributed with parameter a . Figure 6 presents the graphs of the ratio $(1 - H)/(zh)$ and of $1 - 1/E(\tilde{z}|\tilde{z} > z)$ (the later one is plotted in dashed line and is given for reference because it also tends to $1/a$). The hazard ratio $(1 - H)/(zh)$ is noisier than the conditional mean income ratio which is not surprising. Asymptotic values are roughly the same for incomes above \$200,000. Both curves are U-shaped but the pattern of the two curves below the \$200,000 income level are different: the hazard ratio is much higher for low incomes, it decreases faster until income level \$80,000; the hazard ratio then increases faster until \$200,000. From \$80,000 to \$200,000, the hazard ratio increases from 0.32 to 0.55. This pattern suggests that optimal rates should be also U-shaped: high marginal rates for low incomes, decreasing marginal rates until \$80,000 and then increase in marginal rates until level \$200,000. This particular pattern of the hazard ratio confirms the previous intuition that increasing marginal rates at high income levels are justified from an optimal taxation point of view if elasticities are constant.

³¹Only 1991, which was a sharp recession year, is lower.

Estimates of high income optimal tax rates

Table 1 presents optimal asymptotic rates using formula (1.12) for a range of realistic values for the Pareto parameter of the income distribution, $\bar{\zeta}^u$ and $\bar{\zeta}^c$, (the asymptotic elasticities) and \bar{g} (ratio of social marginal utility of income for infinite income to the marginal value of public funds³²). The Pareto parameter takes 3 values: 1.5, 2 and 2.5. Empirical wage distributions have a Pareto parameter close to 2 and AGI distributions have a parameter closer to 1.5. In the 1960s and 1970s the Pareto parameter of wages and AGI distributions were slightly higher (around 2.5). Uncompensated elasticity takes three values: 0, 0.2 and 0.5. Compensated elasticity takes 3 values: 0.2, 0.5 and 0.8. Two values are chosen for \bar{g} : 0 and 0.25.

Except in the cases of high elasticities, the optimal rates are fairly high. Comparing the rows in Table 1, it appears that the Pareto parameter has a big impact on the optimal rate. Comparing columns (2), (5) and (7) (or columns (3), (6), (8)), we see that at fixed compensated elasticity, the optimal rate is very sensitive to the uncompensated elasticity. This confirms the intuition that deadweight burden computations, which depend only on compensated elasticities, may be misleading when discussing tax reforms.

The most convincing elasticity estimates from the empirical literature suggest that the long-term compensated elasticity should not be bigger than 0.5 and that the uncompensated elasticity is probably even smaller. Table 1 suggests that in this case, the optimal top rate on labor income should not be lower than 50% and maybe as high as 80%.

1.5.2 Numerical simulations of tax schedules

I now present simulations using the distribution of wages of 1992. I use utility functions with constant compensated elasticity ζ^c . Fixing the compensated elasticity has several advantages. First, the compensated elasticity is the key parameter of most empirical studies and therefore, having this parameter fixed over the whole popula-

³²Diamond (1998) presented a table of asymptotic rates in function of the Pareto parameter a of the *skill* distribution, the elasticity of earnings (in the case he considers, compensated and uncompensated elasticities are identical) and the ratio \bar{g} . He looked at a wider range of Pareto parameters but confused a and $1 + a$ in selecting examples.

tion provides a good benchmark for simulations. Second, deadweight burdens are very easy to compute for utility functions with constant compensated elasticity functions. I derive in appendix the general form of utility functions with constant compensated elasticity. In the simulations, I use two types of utility functions with constant elasticities.

With utility functions of Type I, there are no income effects and therefore compensated and uncompensated elasticities are the same. The utility function takes to following form:

$$u = \log\left(c - \frac{l^{1+k}}{1+k}\right) \quad (1.36)$$

The elasticity is equal to $1/k$. This case was examined by Atkinson (1990) and Diamond (1998). Maximization of this utility function with a linear budget constraint $c = n(1-\tau) + R$ leads to the following first order condition: $l = (n(1-\tau))^\zeta$. Therefore, labor supply l tends to infinity at rate n^ζ . Moreover, positive tax rates reduce labor supply by a factor $(1-\tau)^\zeta$ and therefore have a large negative impact on output.

Type II utility functions are such that,

$$u = \log(c) - \log\left(1 + \frac{l^{1+k}}{1+k}\right) \quad (1.37)$$

The compensated elasticity is equal to $1/k$ but there are income effects. The uncompensated elasticity ζ^u can be shown to tend to zero when n tends to infinity. Realistically, when n increases to infinity, l can be shown to tend to a finite limit equal to $\bar{l} = [(1+k)/k]^{1/(k+1)}$ whatever the linear tax rate τ is. Therefore, taxes have not such a negative impact on output compared to the previous utility function.

I use the wage income distribution of year 1992 to perform numerical simulations. The skill distribution is calibrated such that given the utility function and the *actual* tax schedule, the resulting income distribution replicates the *empirical* wage income distribution. The original Mirrlees (1971) method of computation will be used. The main difficulty here comes from the fact that the empirical distribution is used. The details of the numerical computations are presented in appendix.

Optimal rates are computed such that the ratio of government spending E to aggregate production is equal to 0.25. Optimal rates simulations are performed for

the two types of utility functions, two different social welfare criteria (Utilitarian and Rawlsian) and two compensated elasticity parameters ($\zeta^c = 0.25$ and $\zeta^c = 0.5$). Because for both types of utility functions, $u_c \rightarrow 0$ as $n \rightarrow \infty$, \bar{g} is always equal to zero and thus the asymptotic rates are the same with both welfare criteria. The social marginal weights $MS(z)$ are roughly decreasing at the rate $1/z$.

Results are reported on Figures 7 to 10. Optimal marginal rates are plotted for yearly wage incomes between 0 and 300,000 dollars. The curves represent the optimal non-linear marginal rates and the dotted horizontal lines represent the optimal linear rates (see below). As expected, the precise level of the rates depends on the elasticities and on the type of the utility function. In all cases, however, the optimal rates are clearly U-shaped.³³ Optimal rates are decreasing from \$10,000 to \$75,000 and then increase until income level \$200,000. Above \$200,000 the optimal rates are close to their asymptotic level.

As expected, the Rawlsian criterion leads to higher marginal rates (note that Rawlsian marginal rates at the bottom are equal to one). The difference in rates between the two welfare criteria is larger at low incomes and decreases smoothly toward 0 (the asymptotic rates are the same). As a consequence, the U-shape is less pronounced for the Rawlsian criterion than for the Utilitarian criterion (compare Figures 7 and 9 and Figures 8 and 10). Unsurprisingly, higher elasticities lead to lower marginal rates. Note also that higher elasticities imply a more pronounced U-shape and therefore a more non-linear tax schedule.

I have also reported on the Figures the optimal *linear* rates computed for the same utility functions, welfare criteria and skill distribution. The optimal linear rates are also computed so that government spending over production be equal to 0.25. The optimal rates are represented by the horizontal dotted lines (the upper one corresponding to $\zeta^c = 0.25$ and the lower one to $\zeta^c = 0.5$). Table 2 reports the optimal average rates³⁴ in the non-linear case along with the optimal linear rate.³⁵ The guaranteed consumption levels of people with skill zero (who supply zero labor and thus earn zero income) in terms of average income are also reported. As average

³³The rate at the bottom is not zero because labor supply tends to zero as the skill n tends to zero, violating one of the assumptions of Seade (1977).

³⁴The average is weighted by incomes (i.e. $\int zT'(z)h(z)dz / \int zh(z)dz$).

³⁵The asymptotic rate in the non-linear case is reported in parenthesis.

incomes differ in the linear and non-linear cases, I report (in parenthesis), below the guaranteed income level for the linear case, the ratio of the guaranteed income for the linear case to the guaranteed income for the non-linear case: this ratio allows a simple comparison between the absolute levels of consumption of the least skilled individuals in the linear and non-linear case.

The average marginal rates are lower in the non-linear cases than in the linear cases. The guaranteed levels of consumption are slightly higher in relative terms in the linear cases (than in the non-linear cases) but as production is lower in the linear cases, the absolute levels are similar. Therefore, non-linear taxation is significantly more efficient than linear taxation to redistribute income. In particular, it is better from an efficiency point of view to have high marginal rates at the bottom (which corresponds to the phasing out of the guaranteed income level). It should be noted also that the linear rate is higher than the non-linear asymptotic rate in the Rawlsian case but the reverse is true in the utilitarian case. With a utilitarian criterion, high income earners face higher marginal tax rates (and therefore end up paying more taxes) in the non-linear case than in the linear case.

Mirrlees (1971) found much smaller optimal marginal rates in the simulations he presented. Rates were slightly decreasing along the income distribution and around 20% to 30%. The smaller rates he found were the consequence of two effects. First, the utility function he chose ($u = \log(c) + \log(1 - l)$) implies high elasticities. Income effects are constant with $mpe = -0.5$ and compensated elasticities are large with ζ^c decreasing from around 1 (at the bottom decile) to 0.5 (at the top decile). These high elasticities lead to low optimal tax rates. Second, the log-normal distribution for skills implies that the hazard ratio $(1 - H(z))/(zh(z))$ is decreasing over the income distribution and tends to zero as income tends to infinity. This implied a decreasing pattern of optimal rates.

Subsequently, Tuomala (1990) presented simulations of optimal rates using utility functions with smaller elasticities.³⁶ Unsurprisingly, he found higher tax rates but

³⁶As in Stern (1976) for the linear tax case, Tuomala (1990) used the concept of elasticity of substitution between consumption and leisure to calibrate utility functions. This concept does not map in any simple way into the concepts of income effects and elasticities used in the present paper. Tuomala's utility function implies that compensated elasticity are around 0.5 but income effects are large ($mpe \simeq -1$) implying negative uncompensated elasticities.

because he still used a log-normal distribution of skills. The pattern of optimal rates was still regressive, from around 60% at the bottom to around 25% at 99th percentile. Calibrating carefully the skill distribution on the empirical income distribution is thus of much importance to obtain reliable results with numerical simulations. In particular, using log-normal skill distribution always leads to regressive tax schedules—especially at the high end of income distribution.

1.6 Conclusion

This paper has made an attempt to understand optimal taxation of income using the concepts of compensated and uncompensated elasticities of labor income with respect to marginal tax rates. This approach has proved fruitful on various grounds.

First, a simple formula for optimal asymptotic rates has been derived depending on four key parameters: the compensated and uncompensated average elasticities of high income earners, the conditional mean income ratio (which is the ratio of the mean of incomes above a given level to this level of income), and the redistributive tastes of the government.

The empirical literature on the behavioral effects of taxation has failed to generate a consensus on the size of the elasticities of labor supply. The conditional mean income ratio is much easier to estimate; because of its importance for optimal taxation, this ratio deserves further and more extensive investigation. Empirical distributions of income show that this parameter is roughly constant over a very broad range of high incomes. Therefore, the asymptotic formula for marginal rates is much more relevant empirically than the well known zero marginal top rate result holding for bounded distributions of income. Using elasticity estimates from the empirical literature, the formula for asymptotic top rates suggests that marginal rates for labor income should not be lower than 50% and may be as high as 80%.

Second, it has been shown that optimal tax formulas (both linear and non-linear) can be derived without referring to adverse selection theory by just examining the effects of small tax reforms on reported income and welfare. This method has the advantage of showing precisely how the different economic effects (welfare effects, elasticity effects and income effects) come into play and which are the relevant pa-

rameters for optimal taxation. Deriving optimal rates using the original Mirrlees approach gives no hint about the different effects at play and therefore makes the interpretation of the formulas of optimal taxes much more difficult.³⁷ Moreover, the original Mirrlees approach relies heavily on the fact that all individuals differ only through their skills and thus cannot be generalized to a heterogeneous population. The elasticity method used throughout this paper can be extended much more easily to deal with a heterogeneous population of taxpayers: the same formulas apply once elasticities are considered as the average elasticities over the population at given income levels.

Third, the use of elasticity concepts clarifies the relationship between the distribution of skills and the distribution of incomes. In particular, the Pareto parameters of the income distribution and of the skill distribution are linked through the asymptotic uncompensated elasticity. Numerical simulations could therefore be performed using empirical distribution of wages. The simulations showed that a U-shaped pattern for marginal rates may well be optimal. Marginal rates should be high at low income levels, decrease until the middle class is reached and then increase until it converges to the asymptotic level (which is roughly attained at a level of \$250,000 per year for a household).

My analysis can be extended in a number of ways. First, empirical income distributions deserve further examination. The hazard ratio $(1 - H(z))/(zh(z))$ and the conditional mean income ratio $E(z|z > \bar{z})/\bar{z}$ are particularly interesting because optimal rates are closely related to these ratios. They could be compared across countries and over years. Second, the general framework under which the approach used here to derive optimal tax rates is valid, needs still to be worked out precisely. In particular, knowing whether formula (1.30) could be implemented using a convenient algorithm would be interesting. This would allow for the estimation of optimal non-linear rates without relying on the specific framework of the Mirrlees model. Last, it might be fruitful to apply the same methodology to other tax and redistribution problems. In particular, the issue of optimal tax rates at the bottom of income distribution deserves more attention in order to cast light on the important problem of designing

³⁷This may explain why the theory of optimal income taxation has remained almost ignored by the applied literature in public economics.

income maintenance programs.

◊

Appendix A: Proofs of the Results of Section 4

Derivation of the Mirrlees's FOC for optimal rates (1.21)

Recall that c is defined implicitly as a function of u and l by $u = u(c, l)$. Therefore, $\partial c/\partial u = 1/u_c$ and $\partial c/\partial l = -u_l/u_c$. The first order conditions for the maximization of the Hamiltonian are given by:

$$\dot{\phi}(n) = -\frac{\partial H}{\partial u} = -[G'(u_n) - \frac{p}{u_c^{(n)}}]f(n) + \phi(n)\frac{l_n u_{cl}^{(n)}}{n u_c^{(n)}} \quad (1.38)$$

l_n is chosen so as to maximize H :

$$0 = \frac{\partial H}{\partial l} = p[n + \frac{u_l^{(n)}}{u_c^{(n)}}]f(n) + \phi(n)\frac{\psi_l^{(n)}}{n} \quad (1.39)$$

Equation (1.38) is a linear differential equation in $\phi(n)$ which can be integrated using the standard method and the transversality conditions $\phi(0) = \phi(\infty) = 0$:

$$\phi(n) = -\int_n^\infty [\frac{p}{u_c^{(m)}} - G'(u_m)] \exp[-\int_n^m \frac{l_s u_{cl}^{(s)}}{s u_c^{(s)}} ds] f(m) dm$$

Replacing the integrated expression of $\phi(n)$ into (1.39) gives immediately (1.21). \mathcal{QED}

Proof of Lemma 1

$\dot{z}_n/z_n = (l_n + n\dot{l}_n)/(nl_n)$ and $l_n = l(w_n, R_n)$ where $w_n = n(1 - T')$ is the net-of-tax wage rate and $R_n = nl_n - T(nl_n) - nl_n(1 - T')$ is the virtual income of an individual with skill n . $l(w, R)$ is the uncompensated labor supply function introduced in Section 3. Therefore,

$$\dot{l}_n = \frac{\partial l}{\partial w} [1 - T' - n(n\dot{l}_n + l_n)T''] + \frac{\partial l}{\partial R} (n\dot{l}_n + l_n)(nl_n T'')$$

and rearranging,

$$\dot{l}_n = \frac{w_n}{l} \frac{\partial l}{\partial w} \frac{l}{n} + [w_n \frac{\partial l}{\partial R} - \frac{w_n}{l} \frac{\partial l}{\partial w}] \frac{nl_n T''}{n(1 - T')} [l_n + n\dot{l}_n]$$

Using the definitions (1.1) and (1.2) along with the Slutsky equation (1.4), I obtain:

$$\dot{l}_n = \zeta^u \frac{l_n}{n} - \dot{z}_n \frac{l_n T''}{1 - T'} \zeta^c$$

and therefore,

$$\frac{\dot{z}_n}{z_n} = \frac{n\dot{l}_n + l_n}{nl_n} = \frac{1 + \zeta^u}{n} - \dot{z}_n \frac{T''}{1 - T'} \zeta^c$$

which is exactly (1.22). The second order condition for individual maximization is $\dot{z}_n \geq 0$. Therefore, if (1.22) leads to $\dot{z}_n < 0$, this means that T' decreases too fast producing a discontinuity in the income distribution. *QED*

Proof of Proposition 1

In order to express optimal marginal rates in function of elasticities, I first derive formulas for ζ^u , ζ^c and mpe as a function of the utility function u and its derivatives. The uncompensated labor supply $l(w, R)$ is derived implicitly from the first order condition of the individual maximization program: $wu_c + u_l$. Differentiating this equation with respect to l, w and R leads to:

$$[u_{cc}w^2 + 2u_{cl}w + u_{ll}]dl + [u_c + u_{cc}wl + u_{cl}]dw + [u_{cc}w + u_{lc}]dR = 0$$

Replacing w by $-u_l/u_c$, the following formulas for ζ^u and mpe are obtained:

$$\zeta^u = \frac{u_l/l - (u_l/u_c)^2 u_{cc} + (u_l/u_c) u_{cl}}{u_{ll} + (u_l/u_c)^2 u_{cc} - 2(u_l/u_c) u_{cl}} \quad (1.40)$$

$$mpe = \frac{-(u_l/u_c)^2 u_{cc} + (u_l/u_c) u_{cl}}{u_{ll} + (u_l/u_c)^2 u_{cc} - 2(u_l/u_c) u_{cl}}$$

and using the Slutsky equation (6),

$$\zeta^c = \frac{u_l/l}{u_{ll} + (u_l/u_c)^2 u_{cc} - 2(u_l/u_c) u_{cl}} \quad (1.41)$$

The first order condition of the individual (1.18) leads to $n + u_l/u_c = nT' = -(u_l/u_c)T'/(1 - T')$. Therefore (1.21) can first be rewritten as follows:

$$\frac{T'}{1 - T'} = -\frac{\psi_l}{u_l} \left(\frac{1 - F(n)}{nf(n)} \right) \int_n^\infty \left[1 - \frac{G'(u_m)u_c^{(m)}}{p} \right] \frac{u_c^{(n)}}{u_c^{(m)}} T_{nm} \left(\frac{f(m)}{1 - F(n)} \right) dm \quad (1.42)$$

The first part of (1.42) is equal to $A(n)$ iff $-\psi_l/u_l = (1 + \zeta^u)/\zeta^c$. ψ is defined such

that $\psi(u, l) = -lu_l(c, l)$ where c is a function of (c, l) such that $u = u(c, l)$. Therefore:

$$\psi_l = -u_l - lu_{ll} - lu_{cl} \frac{\partial c}{\partial l} = -u_l - lu_{ll} + lu_{cl} \frac{u_l}{u_c}$$

Now using (1.40) and (1.41), it is easy to see that: $(1 + \zeta^u)/\zeta^c = 1 + lu_{ll}/u_l - lu_{cl}/u_c$ and therefore indeed $-\psi_l/u_l = (1 + \zeta^u)/\zeta^c$.

The second part of (1.42) is equal to $B(n)$ if it is shown that:

$$T_{nm} \frac{u_c^{(n)}}{u_c^{(m)}} = \exp\left[\int_n^m \left(1 - \frac{\zeta_{(s)}^u}{\zeta_{(s)}^c}\right) \frac{\dot{z}_s}{z_s} ds\right]$$

By definition of T_{nm} and expressing $u_c^{(n)}/u_c^{(m)}$ as an integral:

$$T_{nm} u_c^{(n)}/u_c^{(m)} = \exp\left[\int_n^m \left(-\frac{d \log(u_c^{(s)})}{ds} - \frac{l_s u_{cl}^{(s)}}{s u_c^{(s)}}\right) ds\right] \quad (1.43)$$

I note $H(s) = -(du_c^{(s)}/ds + l_s u_{cl}^{(s)}/s)/u_c^{(s)}$ the expression in (1.43) inside the integral.

Now, $u_c^{(s)} = u_c(c_s, l_s)$, therefore

$$du_c^{(s)}/ds = u_{cc}^{(s)} \dot{c}_s + u_{cl}^{(s)} \dot{l}_s \quad (1.44)$$

From (1.19),

$$u_c^{(s)} \dot{c}_s + u_l^{(s)} \dot{l}_s = \dot{u}_s = -l_s u_l^{(s)}/s \quad (1.45)$$

Substituting \dot{c}_s from (1.45) into (1.44), I obtain:

$$du_c^{(s)}/ds = -[s \dot{l}_s + l_s] u_l u_{cc} / (s u_c) + u_{cl} \dot{l}_s$$

Substituting this expression for $du_c^{(s)}/ds$ in $H(s)$ and using again the expressions (1.40), (1.41), we have finally:

$$H(s) = [lu_l u_{cc}/u_c^2 - lu_{cl}/u_c] \left(\frac{l_s + s \dot{l}_s}{s l_s}\right) = \left(\frac{\zeta^c - \zeta^u}{\zeta^c}\right) \frac{\dot{z}_s}{z_s}$$

which finishes the proof. Note that on bunching intervals included in (n, m) , $\dot{z}_s = \dot{c}_s = 0$, $H(s) = 0$, and all the preceding equations remain true, and thus the proof goes through. \mathcal{QED}

Derivation of the formula for optimal rates (1.30) from formula (1.29)

I note,

$$K(z) = \int_z^\infty -mpe \frac{T'}{1-T'} h^*(z') dz'$$

Equation (1.29) can be considered as a first order differential equation in $K(z)$:

$$K'(\bar{z}) = D(\bar{z})[C(\bar{z}) + K(\bar{z})]$$

where $C(\bar{z}) = \int_{\bar{z}}^\infty [1 - \frac{MS(z)}{p}] h(z) dz$ and $D(\bar{z}) = mpe / (\bar{z}\zeta^c)$. Routine integration using the method of the variation of the constant and taking into account that $K(\infty) = 0$, leads to:

$$K(\bar{z}) = - \int_{\bar{z}}^\infty D(z)C(z) \exp[- \int_{\bar{z}}^z D(z') dz'] dz$$

Integration by parts leads to:

$$K(\bar{z}) = - \int_{\bar{z}}^\infty C'(z) \exp[- \int_{\bar{z}}^z D(z') dz'] dz - C(\bar{z}) \quad (1.46)$$

Differentiation of (1.46) leads directly to (1.30). QED

Proof of the implications between the limiting Pareto definitions: (1.33) and (1.34)

Suppose the density $f(n)$ satisfies (1.33). Consider a small $\epsilon > 0$. Then for n large enough,

$$\frac{-(1+\gamma) - \epsilon}{n} \leq f'(n)/f(n) \leq \frac{-(1+\gamma) + \epsilon}{n} \quad (1.47)$$

Integrating (1.47) from n to m leads to

$$(m/n)^{-(1+\gamma)-\epsilon} \leq f(m)/f(n) \leq (m/n)^{-(1+\gamma)+\epsilon} \quad (1.48)$$

Integration of (1.48) times $m^\alpha/n^{\alpha+1}$ over m from n to ∞ leads to:

$$1/(\gamma - \alpha + \epsilon) \leq \int_n^\infty m^\alpha f(m) dm / [n^{\alpha+1} f(n)] \leq 1/(\gamma - \alpha - \epsilon)$$

which implies the following property for ϵ tending to zero,

$$\int_n^\infty m^\alpha f(m)dm/[n^{\alpha+1}f(n)] = 1/(\gamma - \alpha) \quad (1.49)$$

Formula (1.49) is also true for $\gamma = \infty$. This can be proved in a similar way by considering that $f'(n)/f(n) \leq -A/n$ for arbitrary large values A . Assume that property (1.49) holds for $f(n)$ and $\gamma < \infty$. Then taking ratios for any α and $\alpha = 1$ gives immediately (1.34). If $\gamma = \infty$ then property (1.49) shows that for every α , $\int_n^\infty m^\alpha f(m)dm$ converges and therefore tends to zero as n tends to infinity. Moreover, $\int_n^\infty m^\alpha f(m)dm/[n^\alpha(1 - F(n))] \geq 1$. Therefore $n^\alpha(1 - F(n))$ tends also to zero as n tends to infinity. Integration by parts gives,

$$\int_n^\infty m^\alpha f(m)dm = n^\alpha(1 - F(n)) + \alpha \int_n^\infty m^{\alpha-1}[1 - F(m)]dm \quad (1.50)$$

and therefore, using (1.49) with $\alpha = 0$, I have $1 - F(m) \ll mf(m)$ for m large. Thus, (1.50) implies that $\int_n^\infty m^\alpha f(m)dm/[n^\alpha(1 - F(n))]$ tends to one as n tends to infinity.

QED

Proof of Proposition 2

I note $H(z)$ the distribution function of earnings and $h(z)$ its density function. I want to show that $I = \int_{z_n}^\infty z_m^\alpha h(z_m)dz_m/[z_n^\alpha(1 - H(z_n))]$ converges to $a/(a - \alpha)$ as z_n (or equivalently³⁸ n) tends to infinity. Changing variables and expressing $\log(z_m/z_n)$ as an integral, I can rewrite I as follows:

$$I = \int_n^\infty \exp\left[\alpha \int_n^m \frac{\dot{z}_s}{z_s} ds\right] f(m)dm/[1 - F(n)] \quad (1.51)$$

I assumed in Proposition 2 that there is no bunching nor gaps, thus I can replace \dot{z}_s/z_s in (1.51) using Lemma 1,

$$\int_n^m \frac{\dot{z}_s}{z_s} ds = \int_n^m \left[\frac{1 + \zeta^u}{s} - \dot{z}_s \frac{T''}{1 - T'} \zeta^c \right] ds \quad (1.52)$$

If I assume that the elasticities ζ^c and ζ^u are constant and equal to $\bar{\zeta}^c$ and $\bar{\zeta}^u$ above a given value n , I can compute exactly the integral on the right hand side of (1.52):

³⁸This point is clear by considering equation (1.53) below.

$$\int_n^m \frac{\dot{z}_s}{z_s} ds = (1 + \bar{\zeta}^u) \log\left(\frac{m}{n}\right) + \bar{\zeta}^c \log\left(\frac{1 - T'_m}{1 - T'_n}\right) \quad (1.53)$$

Therefore,

$$I = \int_n^\infty \left(\frac{m}{n}\right)^{\alpha(1+\bar{\zeta}^u)} \left(\frac{1 - T'_m}{1 - T'_n}\right)^{\alpha\bar{\zeta}^c} \frac{f(m)}{1 - F(n)} dm \quad (1.54)$$

Now using the assumption that T' converges to $\bar{\tau}$ (remember $\bar{\tau} < 1$), and using property (1.34) with $\alpha(1 + \bar{\zeta}^u)$, I have:

$$\lim_{n \rightarrow \infty} I = \frac{\gamma}{\gamma - \alpha(1 + \bar{\zeta}^u)} \quad (1.55)$$

which shows that $H(z)$ satisfies the weak Pareto property (1.34) with parameter $a = \gamma/(1 + \bar{\zeta}^u)$.

The proof assuming only that elasticities converge is similar but more technical. I can compute the two terms of the integral in (1.52) only approximately. Let ϵ be any (small) positive number. For n large enough and any $s \geq n$ (remember $\bar{\zeta}^u > -1$),

$$0 < (1 + \bar{\zeta}^u - \epsilon)/s \leq (1 + \zeta^u)/s \leq (1 + \bar{\zeta}^u + \epsilon)/s \quad (1.56)$$

Integrating (1.56) from n to m , I obtain,

$$\left(\frac{m}{n}\right)^{\alpha(1+\bar{\zeta}^u-\epsilon)} \leq \exp\left[\int_n^m \frac{\alpha(1+\zeta^u)}{s} ds\right] \leq \left(\frac{m}{n}\right)^{\alpha(1+\bar{\zeta}^u+\epsilon)} \quad (1.57)$$

The second term in (1.52) is harder to control. I note $\delta(s) = -\dot{z}_s T''/(1 - T')$. I have used above that $\int_n^m \delta(s) ds = \log[(1 - T'_m)/(1 - T'_n)]$ tends to zero as n tends to infinity. The proof is already done in the case ζ^c constant, therefore my goal is to bound $D = \int_n^m \zeta^c \delta(s) ds - \bar{\zeta}^c \int_n^m \delta(s) ds$.

- First case: $\bar{\zeta}^c > 0$

Because z_n is increasing, by Lemma 1's result, for n large enough, there is a constant C such that: $\delta(s) \geq -C/s$. Now, I write $\delta(s) = \delta(s)^+ - \delta(s)^-$ where δ^+ and δ^- are the positive and negative³⁹ parts of δ . $0 \leq \delta(s)^- \leq C/s$ and thus $0 \leq \int_n^m \delta(s)^- ds \leq C \log(m/n)$.

³⁹That is, $\delta^+ = \max(0, \delta)$ and $\delta^- = \max(0, -\delta)$ and therefore $|\delta| = \delta^+ + \delta^-$.

As $\int_n^m \delta(s)ds = \log[(1 - T'_{(m)})/(1 - T'_{(n)})]$, I have also, for n large enough,

$$0 \leq \int_n^m \delta(s)^+ ds \leq \log[(1 - T'_{(m)})/(1 - T'_{(n)})] + C \log(m/n) \quad (1.58)$$

Now, for n large enough, because $\zeta^c \rightarrow \bar{\zeta}^c$,

$$|D| \leq \int_n^m |\zeta^c - \bar{\zeta}^c| [\delta(s)^- + \delta(s)^+] ds \leq \epsilon \log(m/n) + \epsilon \log[(1 - T'_{(m)})/(1 - T'_{(n)})] \quad (1.59)$$

which implies for n large enough (remember T' converges to $\bar{\tau} < 1$),

$$-\epsilon \log\left(\frac{m}{n}\right) - \epsilon \leq \int_n^m \zeta^c \delta(s) ds \leq \epsilon \log\left(\frac{m}{n}\right) + \epsilon \quad (1.60)$$

and therefore,

$$(1 - 2\alpha\epsilon) \left(\frac{m}{n}\right)^{-\epsilon\alpha} \leq \exp\left[\int_n^m \alpha \zeta^c \delta(s) ds\right] \leq (1 + 2\alpha\epsilon) \left(\frac{m}{n}\right)^{\epsilon\alpha} \quad (1.61)$$

Multiplying equations (1.57) and (1.61) and integrating over m from n to ∞ leads to result (1.55).

• Second case: $\bar{\zeta}^c = 0$. I assumed in this case that $\zeta^c \downarrow 0$. Using integration by parts,

$$\int_n^m \zeta_{(s)}^c \delta(s) ds = \zeta_{(m)}^c \int_n^m \delta(s) ds - \int_n^m \dot{\zeta}_{(s)}^c \left[\int_n^s \delta(u) du \right] ds \quad (1.62)$$

The first term in (1.62) is clearly converging to 0. Because $\dot{\zeta}_{(s)}^c \leq 0$, the second term can be bounded as follows,

$$\left| \int_n^m \dot{\zeta}_{(s)}^c \left[\int_n^s \delta(u) du \right] ds \right| \leq \int_n^m -\dot{\zeta}_{(s)}^c \left| \int_n^s \delta(u) du \right| ds \leq C \int_n^m -\dot{\zeta}_{(s)}^c ds = C[\zeta_{(n)}^c - \zeta_{(m)}^c]$$

which tends to zero as n tends to infinity. Therefore, in this second case, we have, $\int_n^m \zeta_{(s)}^c \delta(s) ds \rightarrow 0$ and therefore an inequality of the kind of (1.61) can be obtained and the same proof can go through. \mathcal{QED}

Proof of Proposition 3

- Case: $\bar{\zeta}^c = 0$

Because the exponential term inside $B(n)$ is bigger (or equal) to one (see (1.25)) and $G'(u)u_c/p$ tends to $\bar{g} < 1$, for n large enough, $B(n) \geq (1 - \bar{g})/2$. Now because ζ^c tends to zero, $A(n)$ tends to infinity;⁴⁰ therefore T' tends to one.

- Case: $\bar{\zeta}^c > 0$

I assume first that $\gamma < \infty$ and that the formula for $\bar{\tau}$ in Proposition 3 is such that $\bar{\tau} < 1$. Using Lemma 1's result, and noting again $\delta(n) = -\dot{z}_n T''/(1 - T')$,

$$\log(S_{nm}) = \int_n^m \left(1 - \frac{\zeta^u}{\zeta^c}\right) \frac{\dot{z}_s}{z_s} ds = \int_n^m \left(1 - \frac{\zeta^u}{\zeta^c}\right) \left[\frac{1 + \zeta^u}{s} + \zeta^c \delta(s) \right] ds$$

The same computations as in Proposition 2 (case $\bar{\zeta}^c > 0$) lead to:

$$\int_n^m \left(1 - \frac{\zeta^u}{\zeta^c}\right) \zeta^c \delta(s) ds = o(1) \log\left(\frac{m}{n}\right) + [-\overline{mpe} + o(1)] \log\left(\frac{1 - T'_{(m)}}{1 - T'_{(n)}}\right)$$

where $o(1)$ are real functions (of m and n) tending to zero as n tends to infinity.

Now,

$$S_{nm} = \left(\frac{m}{n}\right)^{(1 - \bar{\zeta}^u/\bar{\zeta}^c)(1 + \bar{\zeta}^u) + o(1)} \left(\frac{1 - T'_{(m)}}{1 - T'_{(n)}}\right)^{-\overline{mpe} + o(1)} \quad (1.63)$$

For ease of notation, let $\bar{H} = (1 - \bar{\zeta}^u/\bar{\zeta}^c)(1 + \bar{\zeta}^u)$. Using property (1.33) for $F(n)$, I have (see equations (1.47) and (1.48) and remember $\gamma < \infty$),

$$\frac{f(m)}{f(n)} = \left(\frac{m}{n}\right)^{-(1 + \gamma) + o(1)} \quad (1.64)$$

Now, using (1.63), (1.64) and the expression for T' in Proposition 1,

$$\frac{T'_{(n)}}{1 - T'_{(n)}} = (1 + o(1)) \left(\frac{1 + \bar{\zeta}^u}{\bar{\zeta}^c}\right) (1 - \bar{g}) \int_n^\infty \left(\frac{m}{n}\right)^{\bar{H} - (1 + \gamma) + o(1)} \left(\frac{1 - T'_{(m)}}{1 - T'_{(n)}}\right)^{-\overline{mpe} + o(1)} \frac{dm}{n} \quad (1.65)$$

Routine algebra shows that,

⁴⁰ Assuming $\gamma < \infty$.

$$\frac{\bar{\tau}}{1 - \bar{\tau}} = \left(\frac{1 + \bar{\zeta}^u}{\bar{\zeta}^c} \right) \left(\frac{1 - \bar{g}}{\gamma - \bar{H}} \right)$$

where $\bar{\tau}$ is the expression for the asymptotic optimal rate stated in Proposition 3.⁴¹ Therefore, equation (1.65) can be rewritten as follows:

$$\frac{T'_{(n)}}{(1 - T'_{(n)})^{1-\alpha+o(1)}} = (1 + o(1)) \left(\frac{\bar{\tau}}{1 - \bar{\tau}} \right) \int_n^\infty \beta \left(\frac{m}{n} \right)^{-1-\beta+o(1)} (1 - T'_{(m)})^{\alpha+o(1)} \frac{dm}{n} \quad (1.66)$$

where $\beta = \gamma - \bar{H} > 0$ and $0 \leq \alpha = -\overline{mpe} < 1$. Equation (1.66) implies that (use $1 - T'_{(m)} \leq 1$),

$$\frac{T'_{(n)}}{(1 - T'_{(n)})^{1-\alpha}} \leq \left(\frac{\bar{\tau}}{1 - \bar{\tau}} \right) (1 + o(1))$$

which implies that there exists some (small) $\delta > 0$ such that $T' < 1 - \delta$ for n large enough. Now, using once again equation (1.66) with the inequality $1 - T' > \delta$, we can see that T' is bounded away from 0. Therefore, for some small $\delta > 0$, I have $\delta < T' < 1 - \delta$ for n large enough.

Equation (1.66) is thus bounded (away from 0 and infinity); therefore, I can rewrite (1.66) as follows:

$$\frac{T'_{(n)}}{(1 - T'_{(n)})^{1-\alpha}} = (1 + o(1)) \left(\frac{\bar{\tau}}{1 - \bar{\tau}} \right) \int_n^\infty \beta \left(\frac{m}{n} \right)^{-1-\beta} (1 - T'_{(m)})^\alpha \frac{dm}{n} \quad (1.67)$$

Note that the $o(1)$ terms have been pulled outside the integral.⁴² I introduce now the following function:

$$q(n) = \int_n^\infty \beta \left(\frac{m}{n} \right)^{-1-\beta} \left(\frac{1 - T'_{(m)}}{1 - \bar{\tau}} \right)^\alpha \frac{dm}{n} \quad (1.68)$$

The derivative of $q(n)$ is equal to:

⁴¹Note that the assumption $\bar{\tau} < 1$ is equivalent to $\gamma > \bar{H}$.

⁴²This can be proved by showing that the difference between the left hand sides of equations (1.66) and (1.67) tends to zero as n tends to infinity.

$$\dot{q}(n) = \frac{\beta}{n} \left[q(n) - \left(\frac{1 - T'_{(m)}}{1 - \bar{\tau}} \right)^\alpha \right]$$

Now, introducing the increasing function φ which is the inverse of the increasing function $x \rightarrow x/(1-x)^{1-\alpha}$, equation (1.67) can be rewritten as follows:

$$T'_{(n)} = \varphi[(1 + o(1))\varphi^{-1}(\bar{\tau})q(n)] \quad (1.69)$$

Therefore, $q(n)$ satisfies the following differential equation:

$$\dot{q}(n) = \frac{\beta}{n} \left[q(n) - \frac{1}{(1 - \bar{\tau})^\alpha} \left(1 - \varphi[(1 + o(1))\varphi^{-1}(\bar{\tau})q(n)] \right)^\alpha \right] \quad (1.70)$$

First, note that the function $V(q) = q - (1 - \varphi[\varphi^{-1}(\bar{\tau})q])^\alpha / (1 - \bar{\tau})^\alpha$ is increasing in q (because φ is increasing) and takes value zero at $q = 1$. Therefore if $o(1) \equiv 0$ then $q(n) \equiv 1$ is an unstable equilibrium point of the differential equation (1.70).⁴³ This property is used to show that even with the $o(1)$, term $q(n)$ tends to one as n tends to infinity.

Suppose that $q(n)$ does not converge to one. Then, there is some $\epsilon > 0$ such that for any $N > 0$, there is some $\bar{n} \geq N$ such that either $q(\bar{n}) < 1 - 2\epsilon$ or $q(\bar{n}) > 1 + 2\epsilon$. Consider first N large such that, $|o(1)| < \epsilon$ for all $n \geq N$. Suppose first, that $q(\bar{n}) < 1 - 2\epsilon$ for some $\bar{n} \geq N$ and thus $(1 + o(1))q(\bar{n}) < 1 - \epsilon$.

But now, equation (1.70) and the fact that function $V(q)$ is increasing, implies that there is a small $\mu = -\beta V(1 - \epsilon) > 0$ such that $\dot{q}(\bar{n}) < -\mu/\bar{n}$. Therefore q is decreasing at \bar{n} and in fact $q(n)$ will never get larger than $1 - 2\epsilon$ for any $n \geq \bar{n}$. If it did, q should have increased at some point $n^* \geq \bar{n}$ such that $q(n^*) \leq 1 - 2\epsilon$ (this is clear by considering a graph) which is impossible. Therefore, $\dot{q}(n) < -\mu/n$ for all $n \geq \bar{n}$. This implies that $q(n)$ tends to $-\infty$ which cannot be because $q(n) \geq 0$ by definition (1.68).

In the same way, if $q(\bar{n}) > 1 + 2\epsilon$ for some $\bar{n} \geq N$, then for a small $\mu > 0$ and all $n \geq \bar{n}$, $\dot{q}(n) > \mu/n$, which implies that $q(n)$ tends to ∞ . But because $1 - T' \leq 1$, definition (1.68) of $q(n)$ shows that this cannot be true either. Therefore, $q(n)$ does

⁴³The equilibrium is unstable because if $q(n) > 1$ then $\dot{q}(n) > 0$ and thus $q(n)$ gets further away from the equilibrium. The similar converse property holds if $q(n) < 1$.

converge to 1 as n tends to infinity. Using (1.69), it is clear then that $T' \rightarrow \bar{\tau}$.

The case $\gamma = \infty$ is easier and can be proved by considering that for any large A , $f(m)/f(n) \leq (m/n)^{-1-A}$ (for n large enough). Equation (1.65) then holds as an inequality (with A in place of γ). First, this inequality shows that $T'_{(n)}$ is bounded away from 1; second, the right-hand side is shown to converge to 0 as A increases implying that T' converges to zero.

Let me last show that, in the case $\gamma < \bar{H}$, $T' \rightarrow 1$. Suppose not, then there is some $\epsilon > 0$ and a sequence (n_k) increasing to infinity such that $1 - T'_{(n_k)} > 2\epsilon$ for all k . As $\bar{\zeta}^c > 0$, I can use that, as in the proof of Proposition 2, there is some $C > 0$ such that $\delta(s) \geq -C/s$ (for s large enough); thus (for $n \leq m$ large enough),

$$\frac{1 - T'_{(m)}}{1 - T'_{(n)}} \geq \left(\frac{m}{n}\right)^{-C}$$

Therefore, for all k large enough and any $0 \leq s \leq 1$,

$$1 - T'_{(n_k+s)} \geq (1 - T'_{(n_k)}) \left(\frac{n_k + s}{n_k}\right)^{-C} \geq \epsilon \quad (1.71)$$

Then, inequality (1.71) can be used to get a positive lower bound of $(1 - T'_{(m)})^{-\overline{mpe} + o(1)}$ over an infinity of intervals $(n_k, n_k + 1)$ in the integral appearing in (1.65). This integral is therefore larger than an infinite diverging sum and thus diverges. This implies that $T'_{(n)} = 1$ for n large enough which is a contradiction. *QED*

Appendix B: Technical Results of Section 5

Utility Functions with Constant Compensated Elasticity

Consider a given indifference curve giving utility u . Along this indifference curve, consumption can be considered as a function of labor: $c = c(l, u)$. The budget set is $c = wl + R$. The bundle (c, l) maximizing the agent's utility is such that:

$$\frac{\partial c(l, u)}{\partial l} = w \quad (1.72)$$

The compensated elasticity is constant (denoted again by ζ^c) if, when the wage rate w increases by 1 percent and the consumer stays on the same indifference curve,

the labor supply increases by ζ^c percent. Therefore:

$$\zeta^c \frac{\Delta w}{w} = \frac{\Delta l}{l} \quad (1.73)$$

Now,

$$\Delta w = \frac{\partial c(l + \Delta l, u)}{\partial l} - \frac{\partial c(l, u)}{\partial l} = \frac{\partial^2 c}{\partial l^2} \Delta l \quad (1.74)$$

Therefore, plugging (1.72) and (1.74) into (1.73):

$$\frac{\partial^2 c / \partial l^2}{\partial c / \partial l} = 1 / (\zeta^c l) \quad (1.75)$$

For ease of notation, let $k = 1/\zeta^c$. Now equation (1.75) can be easily twice integrated along indifference curves to get finally:

$$c = A(u) \frac{l^{1+k}}{1+k} + B(u) \quad (1.76)$$

where $A(u)$ and $B(u)$ are the integration constants. Well behaved indifference curves cannot overlap and therefore by considering the case $l = 0$ in the above equation, $B(u)$ must be strictly increasing and non-negative (to rule out negative consumption). By a recardinalization of u , I assume without loss of generality that $B(u) = e^u$ (I impose $u = -\infty$ when $c = 0$). Now, by considering large values of l , to rule out once again overlapping of indifference curves, it must be the case that $A(u)$ is non-decreasing. Therefore equation (1.76) defines an implicit utility function $u = u(c, l)$ because for each non-negative values of (c, l) there is a unique u solving (1.77). Thus the general form of utility functions with constant compensated elasticity is the following:

$$c = A(u) \frac{l^{1+k}}{1+k} + e^u \quad (1.77)$$

where $k = 1/\zeta^c$ and $A(u)$ is non-decreasing. These utility functions are separable in consumption and labor in two cases: either if $A(u)$ is constant or proportional to e^u . The case $A(u)$ constant leads to a utility function with no income effects. These two types of separable utility functions are used in the simulations. *QED*

Numerical Simulations

To simplify computations, I consider the separable form of the utility functions Type I and II. For Type I, $u = c - l^{k+1}/(k+1)$, and $G(u) = \log(u)$ (in the utilitarian case). For Type II, $u = \log(c) - \log[1 + l^{k+1}/(k+1)]$ and $G(u) = u$ (in the utilitarian case).

For both types of utility functions, optimal rates are computed by solving a system of two differential equations in $u(n)$ and $vr(n)$ where u is the utility level and vr is defined such that:

$$vr(n) = \frac{1}{\psi_l^{(n)}} \left(n + \frac{u_l^{(n)}}{u_c^{(n)}} \right) \quad (1.78)$$

Because of separability, $\psi_l = -u_l - lu_{ll}$ and equation (1.21) can be written as:

$$vr(n) = \frac{1}{nf(n)} \int_n^\infty \left[\frac{1}{u_c^{(m)}} - \frac{G'(u_m)}{p} \right] f(m) dm$$

Therefore, the system of differential equations can be written as follows:

$$\dot{vr} = -\frac{vr}{n} \left(1 + \frac{nf'}{f} \right) - \frac{1}{nu_c} + \frac{G'(u)}{pn}$$

$$\dot{u} = -\frac{lu_l}{n}$$

l and c are implicit functions of u and vr (defined by equations (1.78) and $u = u(c, l)$). The system of differential equations used to solve optimal rates depends on $f(n)$ through the expression⁴⁴ $nf'(n)/f(n)$ which is noisy when taken from empirical data. nf'/f is smoothed using Kernel density methods with large bandwidth. $f(n)$ is derived from the empirical distribution of wage income in such a way that the distribution of income $z(n) = nl(n)$ inferred from $f(n)$ with flat taxes (reproducing roughly the real tax schedule) matches the empirical distribution. nf'/f is taken constant above a large income level (above \$1.5 milion) and such that the Pareto parameter of the income distribution be equal to 1.9.⁴⁵

The differential system is solved using numerical integration methods. In the utilitarian case, a value is assumed for p , then $u(0)$ ($vr(0)$) can be computed as a

⁴⁴This expression should be constant in the case of a perfectly Pareto distributed skill density (see (1.31)).

⁴⁵This matches the empirical wage income distribution of year 1992. Moreover, knowing the asymptotic values simplifies considerably the numerical computations.

function of p and $u(0)$) is chosen such that the system converges to the theoretical asymptotic values. p is adapted through trial and error until government surplus over aggregate production is equal to 0.25. In the Rawlsian case, $G' \equiv 0$ and p is not defined, thus a value is assumed for $u(0)$, then $vr(0)$ is chosen such that the solution converges.⁴⁶ $u(0)$ is adapted until government surplus over aggregate production is equal to 0.25. I check that the optimal solutions lead to increasing earnings z_n .

⁴⁶In theory, $vr(0) = \infty$; therefore in the numerical simulation, the lowest skill is taken small but positive so that the initial value of vr be well defined.

Table I: Asymptotic Marginal Rates (Optimal Rates for High Income Earners)

		Uncompensated Elasticity = 0			Uncompensated Elasticity = 0.2			Uncompensated Elasticity = 0.5		
		Compensated Elasticity			Compensated Elasticity			Compensated Elasticity		
		0.2	0.5	0.8	0.2	0.5	0.8	0.5	0.8	
		(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	
Panel A: social marginal utility with infinite income (g) = 0										
Pareto parameter	1.5	91	80	71	77	69	63	57	53	
	2	83	67	56	71	59	50	50	43	
	2.5	77	57	45	67	51	42	44	37	
Panel B: social marginal utility with infinite income (g) = 0.25										
Pareto parameter	1.5	88	75	65	71	63	56	50	45	
	2	80	60	48	65	52	43	43	37	
	2.5	71	50	38	60	44	32	38	31	

Note: g is the ratio of social marginal utility with infinite income over marginal value of public funds. The Pareto parameter of the income distribution takes values 1.5, 2, 2.5. Optimal rates are computed according to formula (16)

Table II: Numerical simulations for Optimal Tax Rates

	Utilitarian Criterion				Rawlsian Criterion			
	Compensated Elasticity		Compensated Elasticity		Compensated Elasticity		Compensated Elasticity	
	0.25	0.5	0.25	0.5	0.25	0.5	0.25	0.5
Non-linear (1)	Linear (2)	Non-linear (3)	Linear (4)	Non-linear (5)	Linear (6)	Non-linear (7)	Linear (8)	
Panel A: Utility Type I								
Optimal Average Rate (Asymptotic Rate)	0.51 (0.68)	0.61	0.38 (0.51)	0.51	0.68 (0.68)	0.80	0.52 (0.51)	0.67
Guaranteed Income Level (linear over non-linear level)	0.33 (1.03)	0.36	0.21	0.26 (1.09)	0.55	0.55 (0.92)	0.42	0.42 (0.87)
Panel B: Utility Type II								
Optimal Average Rate (Asymptotic Rate)	0.59 (0.81)	0.67	0.48 (0.69)	0.60	0.77 (0.81)	0.88	0.65 (0.69)	0.82
Guaranteed Income Level (linear over non-linear level)	0.40 (1.00)	0.42	0.31	0.35 (1.01)	0.60	0.63 (0.92)	0.50	0.57 (0.92)

Note: In the non-linear case, optimal rates are averaged with income weights; asymptotic rates are reported in parenthesis below average rates. The guaranteed income level is expressed in percentage of average income. The ratio of the absolute guaranteed level in the linear case over the absolute guarantee level in the non-linear case is reported in parenthesis.

Figure 1: Tax Reform Decomposition

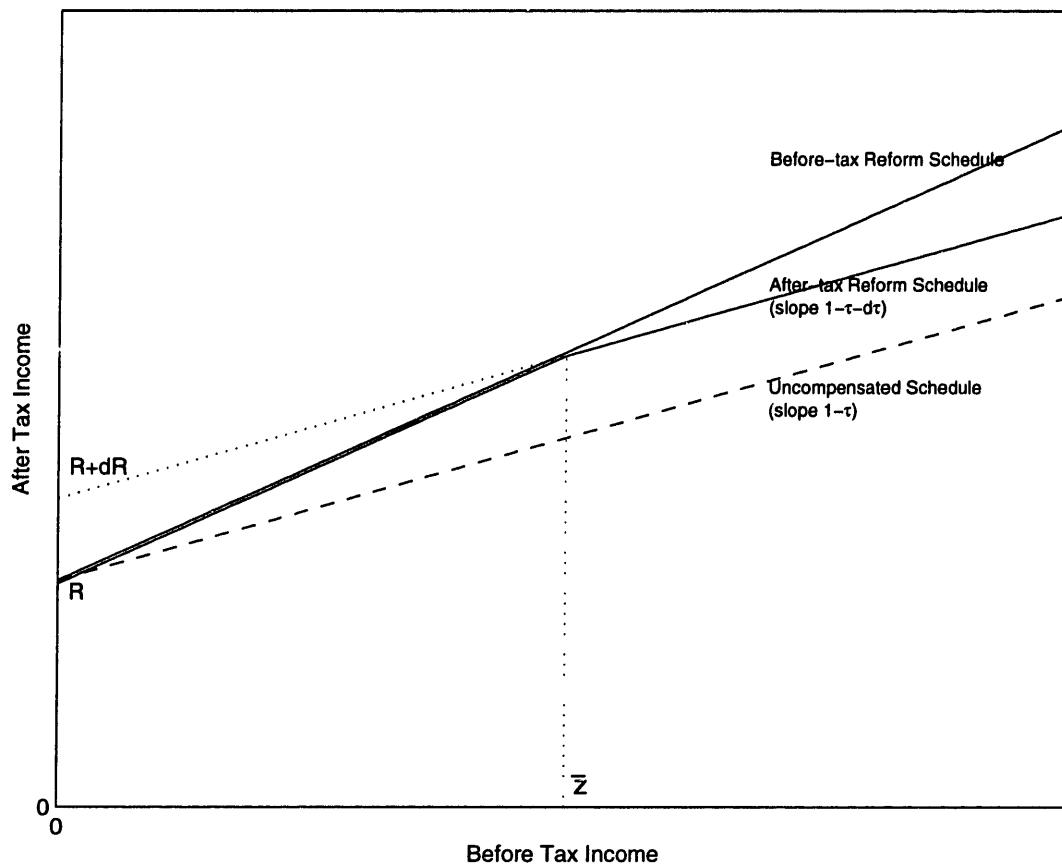


Figure 2: Conditional Mean Income ratios for wages, Years 1992, 1993

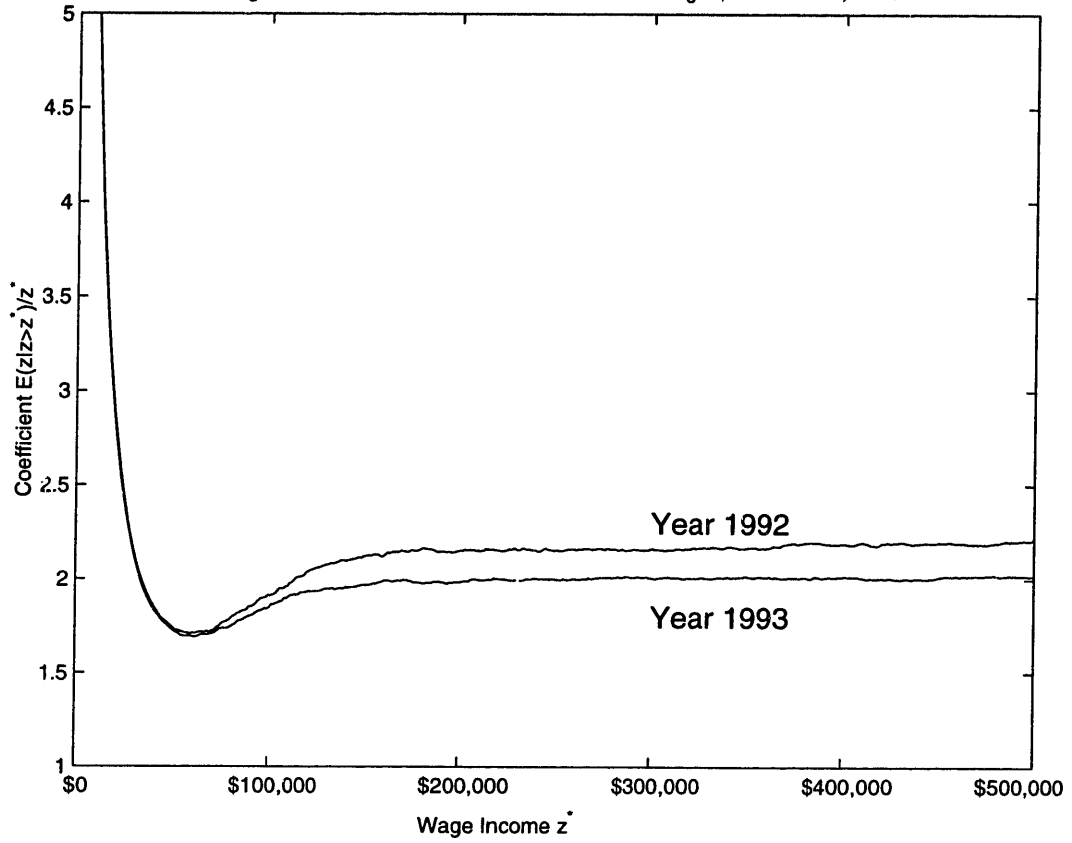


Figure 3: Conditional Mean Income ratios for wages, 1992, 1993 (semilog scale)

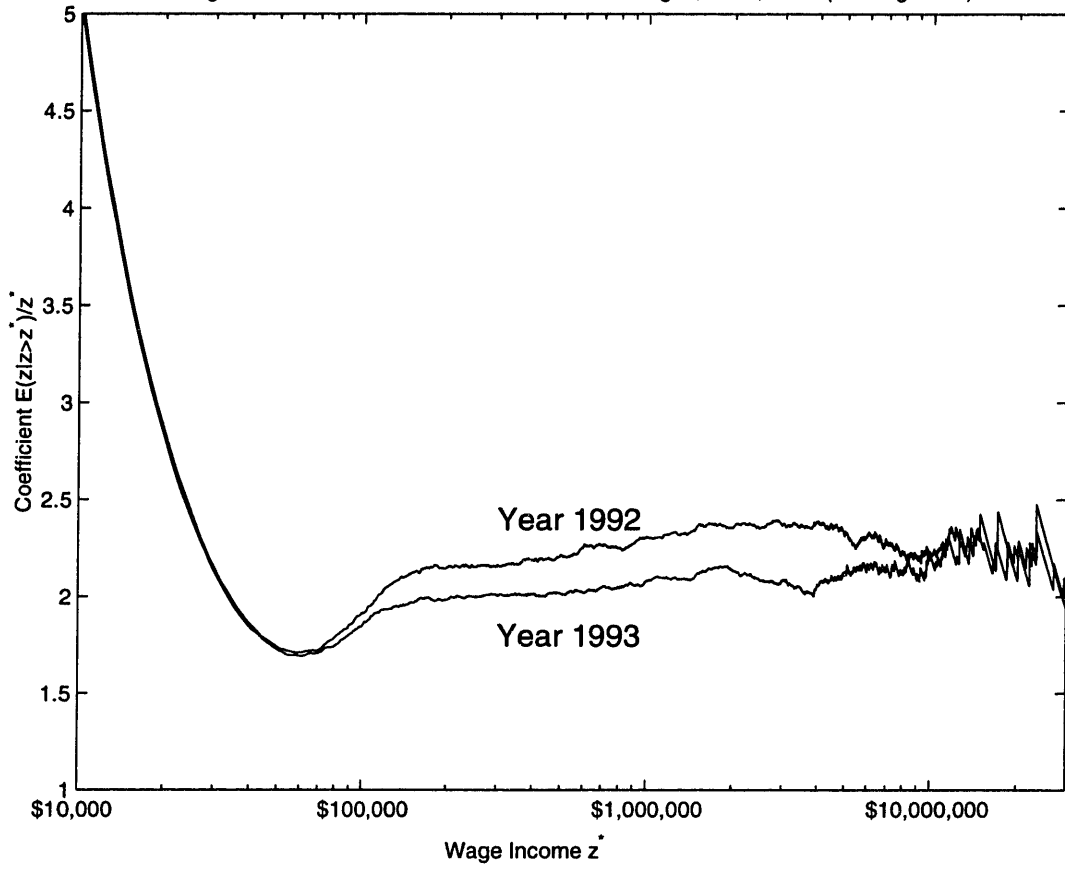


Figure 4: Conditional Mean Income Ratios, Years 1987 to 1993

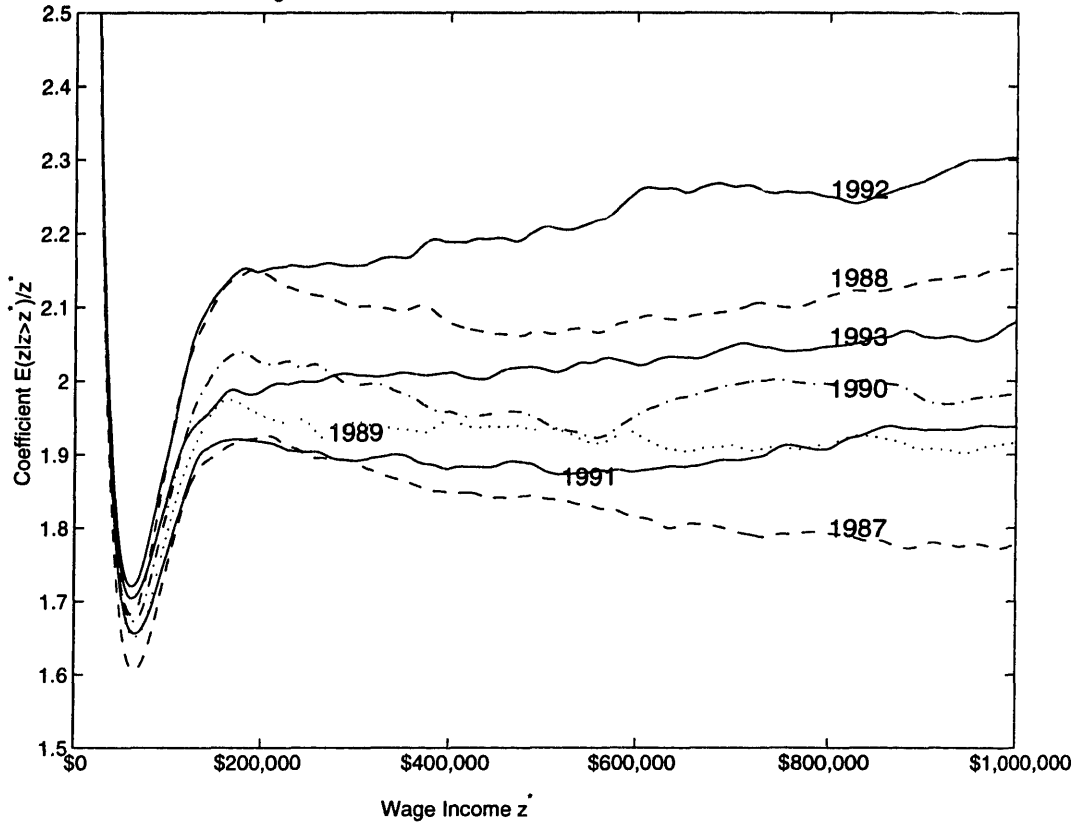


Figure 5: Conditional Mean Income Ratios, Years 1987 to 1993

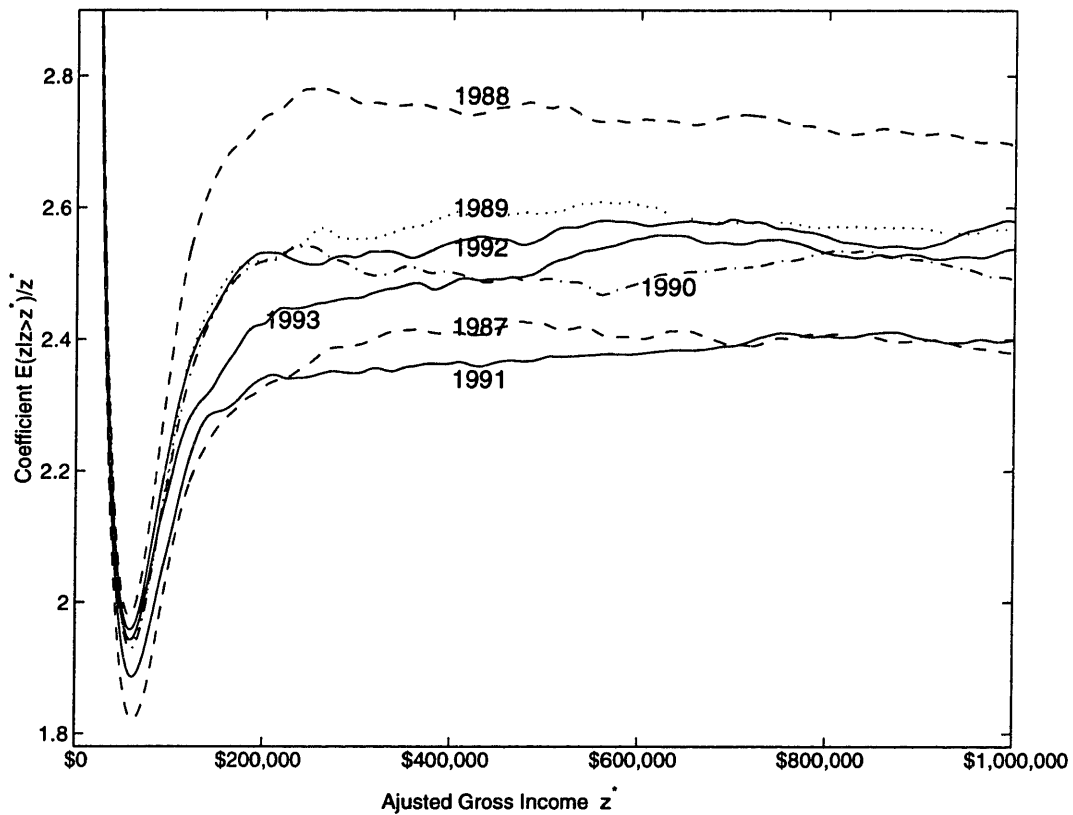


Figure 6: Hazard Ratio $[1-H(z)]/[zh(z)]$ for wages, year 1992

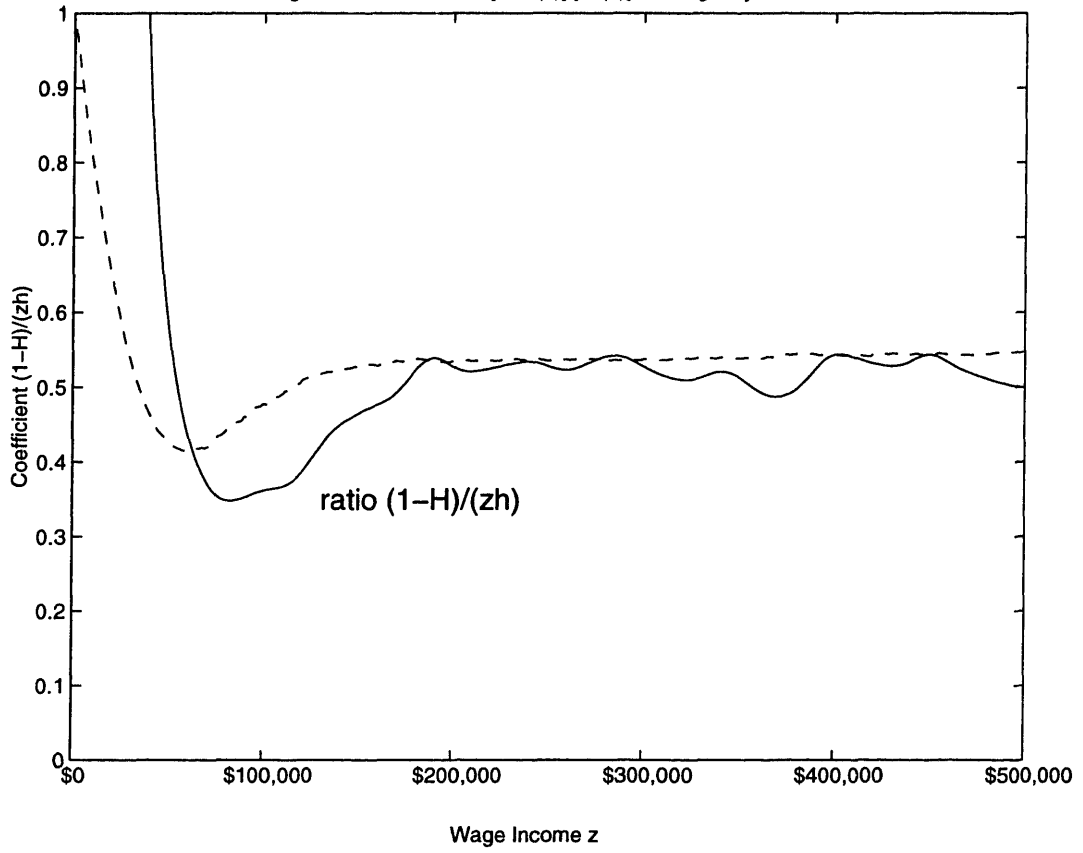


Figure 7: Optimal Rates, Utilitarian Criterion, Utility type I, $\zeta^c=0.25$ and 0.5

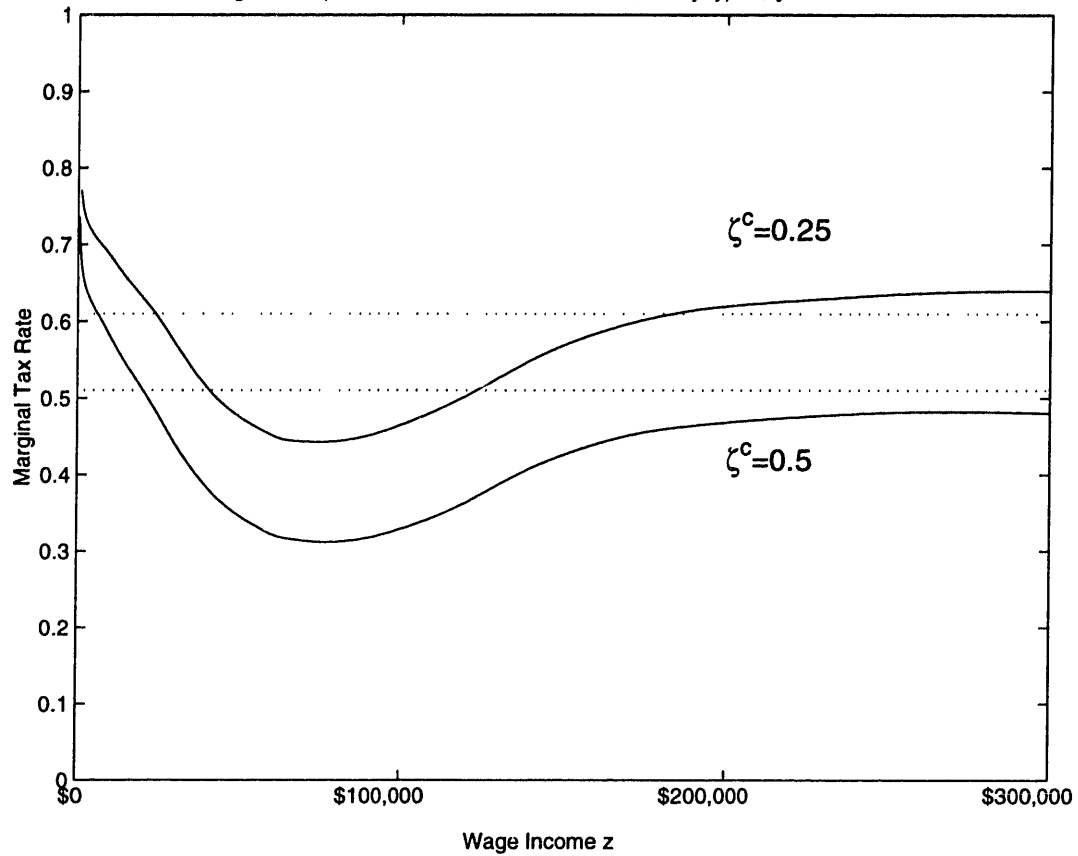


Figure 8: Optimal Rates, Utilitarian Criterion, Utility type II, $\zeta^c=0.25$ and 0.5

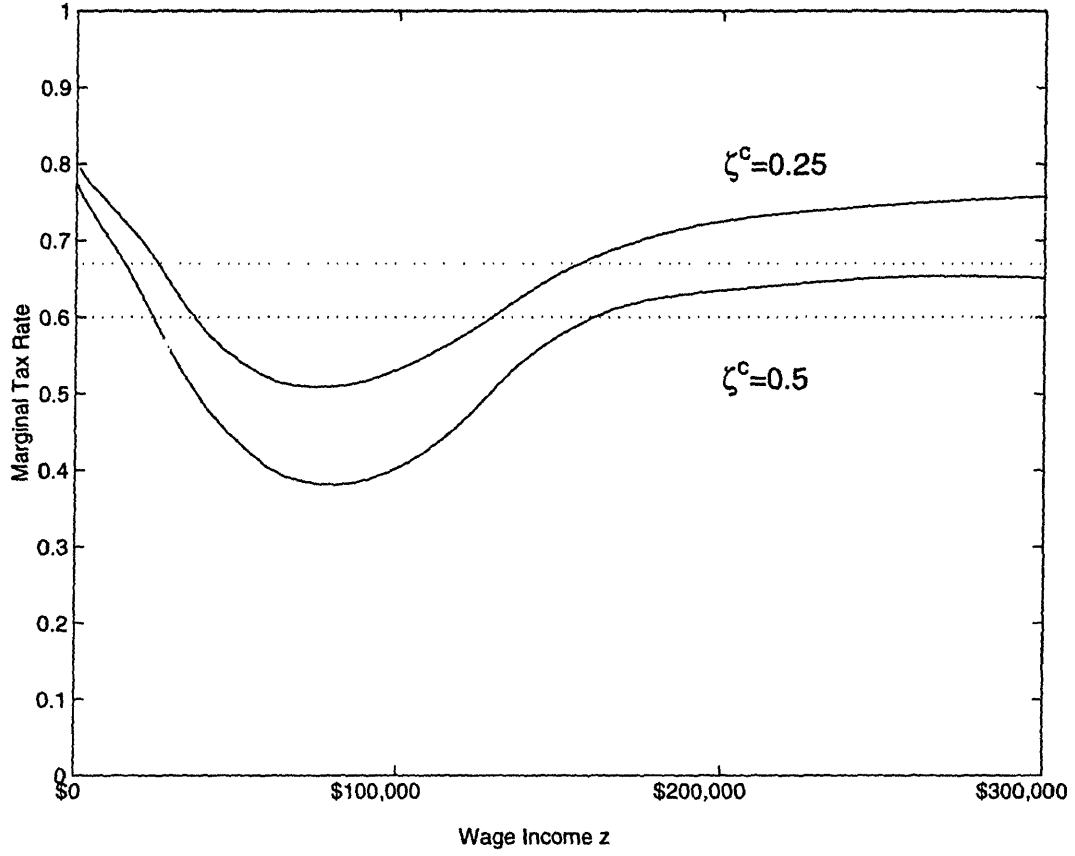


Figure 9: Optimal Rates, Rawlsian Criterion, Utility type I, $\zeta^c=0.25$ and 0.5

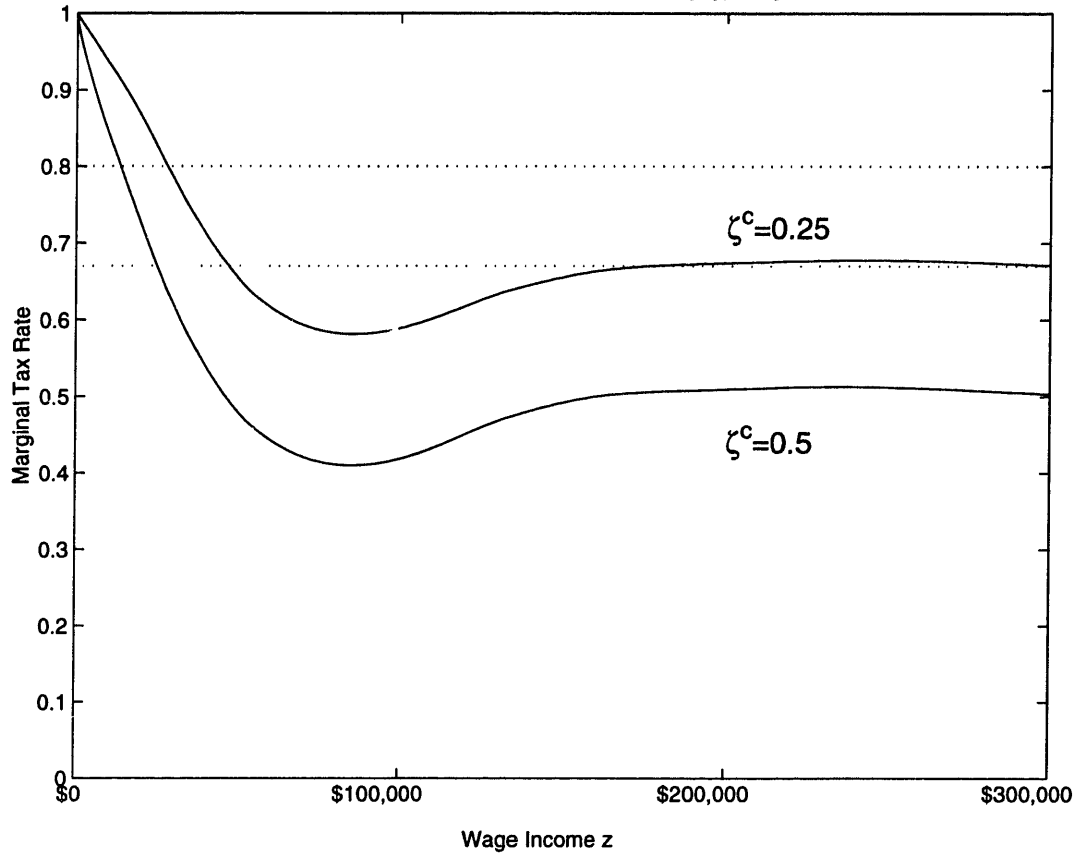
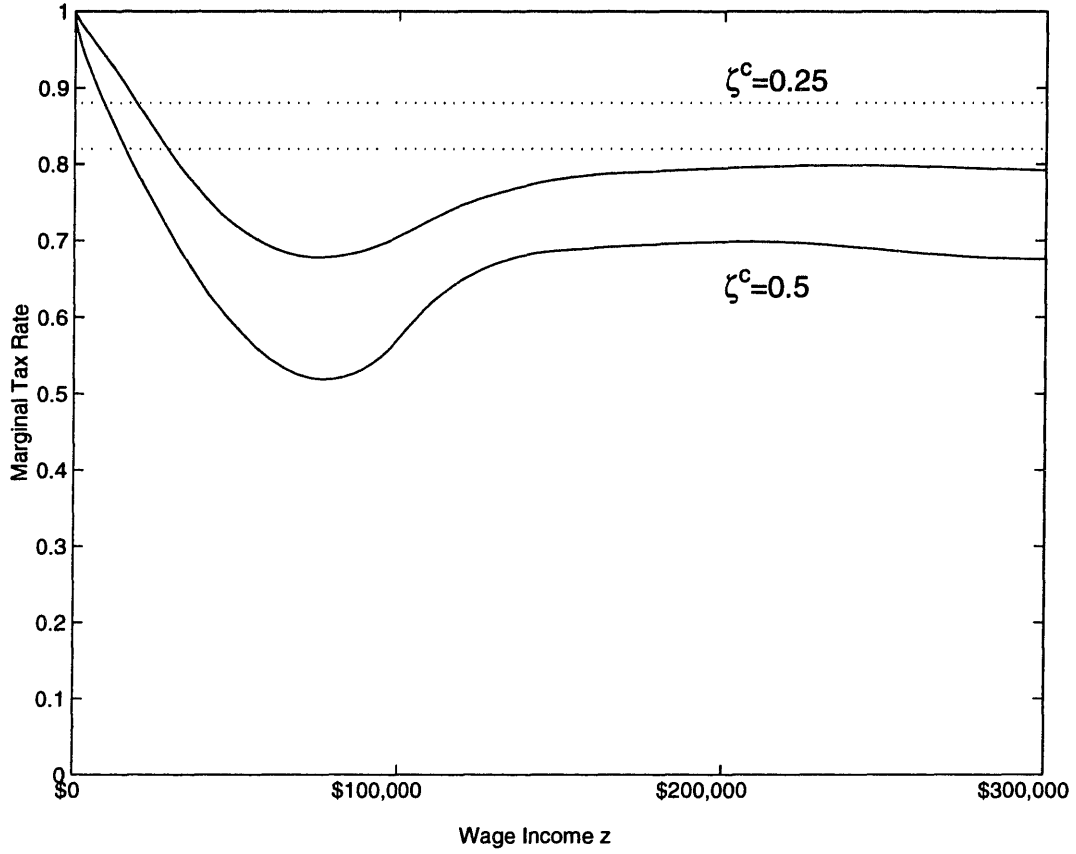


Figure 10: Optimal Rates, Rawlsian Criterion, Utility type II, $\zeta^c=0.25$ and 0.5



Chapter 2

A Characterization of the Income Tax Schedule Minimizing Deadweight Burden

2.1 Introduction

More than seventy years ago, Ramsey (1927) answered the following question. How the government should choose a set of (possibly different) flat tax rates on some or all uses of income so as to raise a given revenue and decrease utility by a minimum? Ramsey voluntarily abstracted from distribution considerations by using a single representative agent model. The present paper asks and answers a similar question for the income tax in a many person two good model (consumption and earnings). The question can be formulated as follows. How the government should set a general income tax (using possibly different marginal tax rates at different income levels) so as to raise a given amount of revenue and reduce utility by a minimum (with no regard for redistribution)? Clearly, the Ramsey representative agent model cannot be used here because heterogeneity in the incomes reported by different taxpayers is needed. As in Ramsey (1927), however, it should be possible to abstract from redistribution concerns by considering that the government gives the same “weight” to every individual in the economy. Surprisingly, this question has never been investigated or even clearly formulated in the public economics literature.

The first issue is to define precisely what is meant by reducing utility by a minimum with no regard for redistribution or giving the same “weight” to everybody. The key tool used in welfare economics to measure welfare loss without taking into account redistributive motives is the deadweight burden concept. The deadweight burden (or loss) of a tax system is the amount lost in excess to what the government collects. The loss arises from the use of distorting rather than non distorting (lump sum) taxes. For a given individual, a positive marginal tax rate on earnings reduces labor supply and thus the government would collect more taxes by using a lump sum tax which keeps the individual on the same indifference curve. This difference in taxes collected is defined as the deadweight burden. For a given marginal tax rate, the higher the substitution effects between leisure and consumption, the higher the distortion and thus the higher is the deadweight burden. The simple approximation, known as Harberger’s triangle formula,¹ which is simply one half of the substitution elasticity times the square of the marginal rate, provides an easy way to evaluate deadweight burdens in many contexts. As the deadweight burden is defined in dollar terms for each individual, the sum of deadweight burdens across individuals is a measure of the total loss due to the tax system which does not discriminate between different individuals: it gives the same “weight” to each individual. Therefore, to answer my initial question, it seems natural to derive the income tax schedule that raises a given amount of revenue and minimizes the sum of deadweight burdens that are incurred by each individual in the economy.

Of course, a trivial solution would be to raise all the revenue required using a poll tax (and applying a zero marginal rate on all income) because such a tax would involve a zero deadweight burden. However, as many taxpayers have earnings potential below the level of the required government spending per capita (even excluding all transfers), this first best tax schedule is not implementable. As some taxpayers have very little earnings potential, the amount of revenue that a government could raise using a poll tax is in fact negligible. In the US in 1996 for example, the government (including state and local taxes) raised about 35% of GDP in taxes. Out of these 35% of GDP, only 17% were transfers. Therefore, even ignoring transfers, the US government has to raise nearly one fifth of total product to finance its needs. This would correspond

¹This approximation formula was first developed in Harberger (1964).

to a \$12,000 poll tax per household! The feasible tax regime which is closest to the poll tax would be to confiscate the incomes of those who cannot pay the poll tax (and impose a zero marginal rate otherwise). This implementable approximation of the poll tax would however impose a 100% tax rate at the bottom of the income distribution. We will see that this policy is far from the second-best optimum. It is unsurprising that in developed countries, governments do not use poll taxes to raise any substantial part of their revenue needs. The only recent exception has been United Kingdom which imposed, from 1988 to 1990, a substantial poll tax to replace local property taxes. The poll tax burden was roughly equal to 400 pounds (around 700 dollars of 1998) per adult person per year. This policy could not be sustained for more than a few years because it appeared rapidly that this tax was imposing too high a burden on poor individuals. Gibson (1990) analyses in detail this episode and even attributes the fall of Mrs Thatcher to the public protest against this tax policy.

Therefore, the poll tax solution can be discarded. Consequently, the initial question asked in this paper is of interest and the features of the solution are not obvious. Most studies using the deadweight burden concept have focused on the comparison of deadweight burden generated by different income tax regimes and thus have in general only provided information about desirable tax reforms. No study has tried to derive directly the general non-linear income tax minimizing distortions. This can be explained by two main reasons. First, the obvious but inadequate poll tax solution may have lead many to think that the problem was either trivial or degenerate. Second, the problem is interesting only for the non-linear tax case because the solution is straightforward in the case of a flat tax: the government would simply set the smallest flat rate which raises the required amount of taxes. Once again, the analogy with the Ramsey problem is useful. The Ramsey problem is interesting because the government can set different tax rates on different commodities.² For the income tax, the problem becomes interesting only if the government can set different marginal rates at different income levels. We will show that, as in the Ramsey case, a uniform rate is not the optimal solution. However, optimizing over any non-linear tax schedules is complicated and requires using mathematical tools more sophisticated than basic

²The primary achievement of Ramsey (1927) was indeed to show that 'the obvious solution that there should be no differentiation is entirely erroneous' (p. 47).

calculus.

Mirrlees (1971) was the first paper which formulated and solved in a rigorous way the general problem of optimal non-linear income taxation with redistributive concerns. The general formula Mirrlees obtained was complicated. Moreover, this formula was expressed in terms of parameters (derivatives of the utility function and skill distribution) which cannot be related in a simple way to empirical magnitudes. In Chapter 1, I have derived directly in terms of substitution and income elasticities and in terms of the income distribution the Mirrlees' general formula for optimal tax rates. This method of derivation displays the key economic effects behind the Mirrlees' formula. Moreover, this alternative method is useful to see how redistributive and efficiency concerns interact. Redistributive concerns are reflected by social marginal weights that the government sets at each income level.

I show in this paper that solving for the tax schedule minimizing deadweight burden is equivalent to a particular Mirrlees optimal income tax problem. I use the same method as in Chapter 1 to derive the tax schedules minimizing deadweight burden because this will allow me to relate the formula to empirical magnitudes. I will discuss how the optimal efficient income tax should look like given elasticities and the actual income distribution. The tax schedule minimizing deadweight burden is an optimal Mirrlees income tax schedule in which the government applies particular marginal welfare weights. Therefore, this paper shows that the Mirrlees framework remains relevant even when the government has no redistributive concerns. In particular the shape of the income distribution is shown to play a key role in the optimal tax schedule minimizing deadweight burden. This income distribution element has not been recognized as crucial by previous studies using the deadweight burden concept because the simple Harberger's triangle formula does not depend on the shape of the income distribution.

In the case of no income effects, the marginal welfare weights are the same for everybody: the tax schedule minimizing deadweight burden is then simply the Mirrlees optimal income tax with no redistributive concerns. This makes intuitive sense, with no redistributive concerns, at the optimum, the government should be indifferent between taking away a small amount from any taxpayer and giving it to any other taxpayer. The features of the solution in the case with no income effects are exactly

equivalent to the classical problem of non-linear pricing of the regulated monopoly. I examine the links between these two problems in detail. In the general case with income effects, however, these marginal welfare weights are higher for people facing higher marginal rates. This last property is examined in detail and interpreted as a deficiency of the deadweight burden concept in the case of income effects. Some studies (Samuelson (1964), Chipman-Moore (1980)) have argued in other contexts why the use deadweight burden is problematic with income effects. The income tax problem I consider gives another illustration of this problem.

The paper is organized as follows. Section 2 presents the deadweight burden concept in the non-linear income tax framework I use. Section 3 solves the optimal efficient tax schedule when there are no income effects and presents a practical implementation of the optimal tax using empirical income distributions. Section 4 analyses the general case with income effects and discusses why the deadweight burden concept is not adequate in this case. Section 5 offers some concluding comments.

2.2 Deadweight Burden Concept

2.2.1 The Model

I consider a standard two good model. A taxpayer maximizes an individual utility function $u = u(c, z)$ which depends positively on consumption c and negatively on earnings z .³ I assume that the individual is on a linear portion of the tax schedule: $c = z(1 - \tau) + R$ where τ is the marginal tax rate at the optimum and R is defined as the virtual income.⁴ The first order condition of this maximization program ($(1 - \tau)u_c + u_z = 0$) defines implicitly a Marshallian (uncompensated) earnings supply function $z = z(1 - \tau, R)$ depending on the marginal tax rate τ and the virtual income R . Similarly, the Hicksian (compensated) earnings function can be defined as the earnings level which minimizes cost $c - z$ needed to reach a given utility level u for a given tax rate τ . I denote it by $z^c = z^c(1 - \tau, u)$. I now recall the definitions

³This framework is a simple extension of the standard labor supply model where utility depends on consumption and labor supply and where earnings is equal to labor supply times an exogenous pre-tax wage rate.

⁴This is the post-tax income the individual would get if he supplied zero labor and was allowed to stay on the “virtual” linear schedule.

of the elasticities and income effects which are used throughout the paper.⁵ From the Marshallian earnings function $z = z(1 - \tau, R)$, the uncompensated elasticity of earnings is defined as:

$$\zeta^u = \frac{1 - \tau}{z} \frac{\partial z}{\partial(1 - \tau)} \quad (2.1)$$

Income effects are defined as:

$$mpe = (1 - \tau) \frac{\partial z}{\partial R} \quad (2.2)$$

If leisure is a non inferior good, then mpe is non-positive. The compensated elasticity of earnings (which is always non-negative) is defined similarly from the Hicksian earnings function $z^c = z^c(1 - \tau, u)$ as:

$$\zeta^c = \frac{1 - \tau}{z} \frac{\partial z^c}{\partial(1 - \tau)} \Big|_u \quad (2.3)$$

Income effects and both elasticities are related through the Slutsky equation:

$$\zeta^c = \zeta^u - mpe \quad (2.4)$$

2.2.2 Deadweight Burden

I present briefly in this section the deadweight burden. Auerbach (1985) provides an extensive presentation of the concept. Figure 1 displays this definition graphically. Assume that the individual faces a general non-linear tax schedule $T(z)$ so that its budget constraint is $c = z - T(z)$. If the tax schedule is well behaved so that the individual optimal choice $(c^* = z^* - T(z^*), z^*)$ satisfies the usual first order condition, then the non-linear budget set can be replaced by a linear budget set characterized by a marginal rate $\tau = T'(z^*)$ and a virtual income $R = z^* - T(z^*) - z^*(1 - \tau)$. Denote by u the utility level $u(c^*, z^*)$ and consider the bundle (c^F, z^F) (F for First-best) that gives utility level u (i.e. $u(c^F, z^F) = u$) but induces no distortions (i.e. the tangent to the indifference curve at this point has slope 1). The deadweight burden (DWB) is the difference between taxes paid at (c^F, z^F) and taxes actually paid at (c^*, z^*) :

⁵Pencavel (1986) provides a comprehensive presentation of these concepts.

$$DWB = z^F - c^F - T(z^*) = z^F(u) - c^F(u) - [z^* - c^*] \quad (2.5)$$

The definition of deadweight burden I presented is based on the equivalent variation. *DWB* is defined by considering the actual second-best bundle (c^*, z^*) and an equivalent (in utility level) first best bundle (c^F, z^F) . An alternative definition of *DWB*, which is sometimes used, is based on the compensating variation. This alternative definition compares the first best bundle with no taxation to the compensated second best bundle (which yields the same utility) but involves the same distortion in prices as the actual bundle (c^*, z^*) . This alternative definition would be less natural for the problem I am considering. I come back to this issue later on.⁶

2.3 Analysis with No Income Effects

2.3.1 Deadweight burden with no income effect

By definition, no income effects means that earnings are not affected by the level of non-labor income ($z = z(1 - \tau, R) = z(1 - \tau)$). Therefore, the indifference curves corresponding to different utility levels are parallel with respect to consumption (vertical axis) as depicted on Figure 2. Therefore, the indifference curves can be represented as $c = u + v(z)$, where u is a utility index, c is a consumption level and z is earned income. The direct utility function can be written (after a suitable cardinalization) $u(c, z) = c - v(z)$. No income effects is thus equivalent to utility functions quasi-linear in consumption.⁷ With no income effects, the Marshallian and Hicksian demand functions are the same and thus uncompensated and compensated elasticities are equivalent. I note ζ this single elasticity. The first order condition for utility maximization (subject to a general non-linear income tax) is:

⁶Kay (1980) shows why the equivalent definition is preferable to the compensating definition in the Ramsey model of commodity taxation. In particular, he shows that the Ramsey optimal tax formulas can be recovered by minimizing deadweight burden based on the equivalent variation. Using the compensating variation does not lead to the same formulas.

⁷Diamond (1998) studied the Mirrlees income tax problem in that particular case with no income effects. I come back to the links between his analysis and the present one later on.

$$1 - T'(z^*) = v'(z^*) \quad (2.6)$$

No income effects simplifies considerably deadweight burden analysis because deadweight burden is not affected by a change in virtual income. The equivalent and compensating variations are the same with no income effects (see Figure 2) and thus the two corresponding definitions of deadweight burden coincide. With the notations of the previous Section and using the quasi-linear property of the utility function, deadweight burden can be written as:

$$DWB = z^F - c^F - (z^* - c^*) = z^F - u - v(z^F) - (z^* - u - v(z^*)) = z^F - v(z^F) - z^* + v(z^*) \quad (2.7)$$

Deadweight burden therefore depends only on earnings z^* and not on the utility level u . Note also that z^F is completely defined by the equation $v'(z^F) = 1$ and thus does not depend on the utility level u . These last two properties are clearly displayed on Figure 2.

2.3.2 Minimization of DWB

I note $H(z)$ the distribution of incomes and $h(z)$ the corresponding density function. The total population is normalized to one. I note p the marginal value of public funds. The formal program is to minimize total deadweight burden (denoted by TDW):

$$TDW = \int DWB(z)h(z)dz$$

Subject to the budget constraint for the government (the multiplier to this constraint is of course p):

$$TTR = \int T(z)h(z)dz \geq E$$

where $DWB(z)$ is the average deadweight burden for taxpayers with income z and E is an exogenous level of taxes that the government must raise in order to finance its needs. In order to rule out the poll-tax solution, consumption is constrained to

be non-negative for each individual. This constraint is kept implicit in the derivation but is examined in detail in the comments of the optimal tax formula. In order to derive a first order condition for optimality, I consider as in Chapter 1 the effect of the following small tax reform: marginal rates are increased by an amount $d\tau$ for incomes between \bar{z} and $\bar{z} + d\bar{z}$. Let us analyse precisely these two effects on total tax receipts and total deadweight burden.

• **Change in Total Tax Receipts**

The tax reform induces no effect on taxpayers with income below \bar{z} . All taxpayers with income above $\bar{z} + d\bar{z}$ pay $d\tau d\bar{z}$ more in taxes. As there are no income effects, their earnings (and thus their deadweight burdens) are not affected. Therefore, the change in tax receipts due to this mechanical effect is simply $[1 - H(\bar{z})]d\tau d\bar{z}$.

The increase $d\tau$ for a taxpayer with income z^* between \bar{z} and $\bar{z} + d\bar{z}$ has an elastic effect which produces a small change in income (denoted by dz^*). Using the definition (2.1) of the elasticity,

$$dz^* = -\zeta_{\bar{z}} \frac{d\tau}{1 - T'} \tag{2.8}$$

As there are $h(\bar{z})d\bar{z}$ taxpayers in the small band $[\bar{z}, \bar{z} + d\bar{z}]$, the total loss in tax receipts due to this elastic effect is,

$$-\zeta_{(\bar{z})} \bar{z} h(\bar{z}) \frac{T'}{1 - T'} d\tau d\bar{z}$$

where $\zeta_{(\bar{z})}$ is the average elasticity at income level \bar{z} .

The overall effect on tax receipts is therefore equal to:

$$dTTR = \left[-\zeta_{(\bar{z})} \bar{z} h(\bar{z}) \frac{T'}{1 - T'} + [1 - H(\bar{z})] \right] d\tau d\bar{z} \tag{2.9}$$

• **Change in Total Deadweight Burden**

As noticed above, because of no income effects, deadweight burden changes only for taxpayers in the band $[\bar{z}, \bar{z} + d\bar{z}]$. A taxpayer with income z^* in the small band changes its earnings by dz^* given by equation (2.8). Using equations (2.6) and (2.7), the change in deadweight burden for this taxpayer is equal to $-[1 - v'(z^*)]dz^* =$

$-T'(z^*)dz^*$. Therefore, using (2.8), the total change in deadweight burden is equal to:

$$dT DW = \left[\zeta_{(\bar{z})} \bar{z} h(\bar{z}) \frac{T'}{1 - T'} \right] d\tau d\bar{z} \quad (2.10)$$

At the optimum, the changes in total deadweight burden ($dTDW$) and in total tax receipts ($dTTR$) must be such that:

$$dT DW = p dTTR \quad (2.11)$$

Using (2.9) and (2.10), we obtain the following formula for the optimal tax rate at income level \bar{z} .

$$\frac{T'(\bar{z})}{1 - T'(\bar{z})} = \frac{1}{\zeta_{(\bar{z})}} \left(\frac{1 - H(\bar{z})}{\bar{z} h(\bar{z})} \right) \left(\frac{p}{p + 1} \right) \quad (2.12)$$

Link with Mirrlees optimal taxation

This derivation shows that solving for the tax schedule minimizing deadweight burden is very close to solving for an optimal tax schedule à la Mirrlees. Comparing the results of Chapter 1 to equation (2.12), it appears clearly that the tax schedule minimizing deadweight burden is in fact an optimal income tax in which the government applies the same welfare weights at all income levels.⁸ The weights are identically equal to $1/(p + 1)$ here.⁹ Therefore, when income effects are negligible and the government minimizes deadweight loss, at the optimum, the government is indifferent between taking a marginal dollar to any taxpayer and redistributing it to any other taxpayer. This exactly amounts to saying that the government has no redistributive motives. Minimizing deadweight burden leads therefore to a sensible solution in the no income effect case.

⁸The welfare weight corresponding to a given income level z represents the social value (expressed in terms of the marginal value of public funds) of an additional dollar of consumption for a taxpayer with income z . With redistributive motives, the welfare weights are typically decreasing with income z .

⁹I come back to the links between minimizing distortions and optimal income taxation in more detail in Section 4.

Comments on Formula (2.12)

The intuition for formula (2.12) is clear. Raising marginal rates at level \bar{z} entails a negative elastic response at this income level which reduces tax receipts and increases excess burden. This behavioral response is proportional to the average elasticity $\zeta_{(\bar{z})}$ at this level and thus $\zeta_{(\bar{z})}$ enters negatively equation (2.12). The negative elastic effect is also proportional to the density of taxpayers $h(\bar{z})$ times the income level \bar{z} . The benefit of this local marginal rate increase is proportional to the number of taxpayers $[1 - H(\bar{z})]$ above income level \bar{z} because increasing marginal rates at \bar{z} increases the tax liabilities of all taxpayers above \bar{z} . These two facts explain the presence of the ratio $(1 - H)/(\bar{z}h)$ in (2.12). This distribution parameter term is not captured by Harberger's triangle formula. The multiplier p represent the cost of raising taxes for public spending. At the optimum, an additional dollar raised in taxes entails an excess burden cost of p dollars. The multiplier p depends positively upon the total amount of taxes E that the government wants to raise.

Formula (2.12) defines marginal tax rates at each income level. As tax liability on income z is the sum of marginal rates from 0 to z plus taxes owed when income is zero ($T(z) = T(0) + \int_0^z T'(s)ds$), the tax schedule will be completely specified once $T(0)$ is defined. If the government could raise all income using a uniform poll tax, it would set $T(0)$ equal to the required amount of taxes per capita. The tax schedule would induce no distortions and thus p and T' would be equal to zero. As discussed in the introduction, the government cannot use an efficient poll tax because low incomes could not pay it. Therefore there is an additional constraint such that $T(0) \leq 0$ because somebody earning nothing cannot presumably pay any taxes. Actually, the constraint should probably be even stronger than that because individuals earning no income would have to be supported (at least partly) by the government. The general constraint on $T(0)$ would thus take the form $T(0) \leq -R$ where R is a non-negative guaranteed income level which everybody is entitled to. On top of this guaranteed income, there is an income tax whose marginal rates are defined by equation (2.12). Now p is chosen such that total taxes collected per capita (net of the guaranteed income R) are exactly equal to the required amount of public spending per capita E . Therefore, changing the level of public spending (or of the guaranteed income level) does affect the level of taxation through the multiplier p but does not affect the global

shape of the marginal tax rate curve.¹⁰

If the government wants to maximize total taxes collected then p is infinite and the term $p/(p + 1)$ in equation (2.12) is then simply equal to one. In that extreme case, the government values public funds infinitely more than private consumption. When p is infinite, formula (2.12) gives therefore the ‘Laffer’ marginal tax rates. The government may want to maximize total tax receipts either because it wants to make public spending E as big as possible or because it wants to make the guaranteed income level as big as possible. In that later case, the government wants to maximize the welfare of the poorest individuals in the economy: this is the maxi-min or Rawlsian welfare criterion.¹¹

The interesting point to note is that the shape of optimal rates is the same when the government uses a Rawlsian criterion (which is the social criterion the most redistributive one can find in the class of Pareto efficient social criteria) and when the government simply minimizes distortions with no concern for redistribution. Therefore, these two criteria, which stand at opposite ends of the political spectrum, lead to the same pattern (but of course not the same level) of marginal rates. The intuition for this somewhat striking result is simple. With the Rawlsian criterion, the government wants to maximize tax receipts but it still has to find the most efficient way to do that.

Link with Ramsey-Boiteux pricing

The derivation presented here is very close to the classical derivation of optimal nonlinear pricing for a regulated monopoly known as Ramsey pricing (see Wilson (1993), Chapters 4 and 5 for a modern exposition). In that model, a monopoly sets a nonlinear price so as to maximize consumer surplus subject to a budget constraint (there are no income effects in the analysis). This is clearly equivalent to the problem considered here. The final pricing formula is expressed as the inverse of the elasticity of demand. More precisely, this demand profile elasticity is defined as the percentage

¹⁰This assertion is only an approximation because changing tax rates can affect the pattern of elasticities $\zeta_{(z)}$ and the shape of the income distribution.

¹¹Atkinson (1975) derived the optimal income tax for the maxi-min criterion using the Mirrlees (1971) method. Recently, Piketty (1997) derived the maxi-min optimal income tax using directly elasticities.

change in the number of customers buying at least a given quantity q which is induced by a one percent change in the marginal price of the q -th unit. It is easy to see how formula (2.12) could also be expressed as a Ramsey inverse elasticity rule:

$$\frac{T'(\bar{z})}{1 - T'(\bar{z})} = \frac{1}{\eta(\bar{z})} \left(\frac{p}{p + 1} \right) \quad (2.13)$$

where,

$$\eta(\bar{z}) = \frac{1 - T'(\bar{z})}{1 - H(\bar{z})} \frac{\partial(1 - H(\bar{z}))}{\partial(1 - T'(\bar{z}))}$$

The elasticity $\eta(\bar{z})$ represents the percentage change in the number of taxpayers with income above \bar{z} when (one minus) the marginal tax rate at \bar{z} is increased by one percent. It is of course possible to show that formulas (2.12) and (2.13) are equivalent using simple calculus. In the income tax case, the supply function with respect to which the elasticity η is taken is not a familiar concept in the public economics field. In the context of the income tax, thinking in terms of labor supply elasticities and the shape of the income distribution makes formula (2.12) easier to analyze and interpret than the equivalent Ramsey inverse elasticity rule (2.13) used in regulation.

Surprisingly, the optimal income tax literature has not explored the link between Ramsey pricing and the Mirrlees model. The models of nonlinear pricing are easier to handle when there are no income effects. The first contributions in the nonlinear pricing literature in Industrial Organization, Mussa-Rosen (1978) and Baron-Myerson (1982), considered the case with no income effects. However, Mirrlees (1971) solved the general case and did not realize that much of the complication was the consequence of income effects. Diamond (1998) was the first to analyze in detail the case with no income effects which simplifies considerably the optimum income tax problem. In summary, there are close links between the early Ramsey (1927) inverse elasticity rule for commodity taxation and the Mirrlees model of nonlinear income taxation. As outlined in the introduction, both problems are conceptually similar. However, solving rigorously the nonlinear income problem required much more complicated mathematical methods and this obscured the conceptual link between these two fundamental tax problems.

2.3.3 Empirical Implementation

The most well known result in non-linear optimal income taxation is that the top marginal rate must be zero if the income distribution is bounded. This result was proved by Sadka (1976) and Seade (1977) and of course goes through in the case considered here (the same proof applies). However, this result is not necessarily true if the income distribution is unbounded. Using a Pareto distribution approximation at the top is empirically relevant over a very broad range of high incomes and thus the tax rate does not converge to zero until very close to the top (see Chapter 1). If the income distribution is Pareto at the top with parameter¹² a then $(1 - H)/(zh) = 1/a$ at the top. Equation (2.12) shows then that optimal rates will not converge to zero (if elasticities do not tend to infinity as income increases). Therefore, the optimal efficient tax rates are very different from the quasi poll-tax solution discussed in the introduction.

The ratio $(1 - H)/(zh)$ has a U-shape pattern as a function of income z . This ratio was studied empirically in Chapter 1. I have reported on Figure 3 this ratio for wage income only for years 1988, 1990 and 1992 using a cross section of tax returns data and including married households only. If elasticities were constant along the income distribution, formula (2.12) would suggest that optimal efficient rates should also be U-shaped.¹³

There is unfortunately no consensus in the empirical literature about the size of the elasticity of income with respect to tax rates. Estimates range very small values (around zero) to very high values (above one). Hausman (1985) and Pencavel (1986) survey the empirical literature on labor supply. However, for a given amount of public spending per capita E , formula (2.12) shows that changing the level of elasticities across all income levels is directly compensated by a change in the multiplier p in order to keep tax receipts constant. Therefore the practical computation of the tax schedule minimizing deadweight burden is not critically sensitive to the assumption about the *absolute* level of elasticities.¹⁴ The pattern of optimal efficient rates is sensitive to the

¹²The density of a Pareto distribution takes the form of a power law: $h(z) = C/z^{a+1}$ where a is the Pareto parameter of the distribution.

¹³This assertion must be taken cautiously because the income distribution and thus the ratio $(1 - H)/(zh)$ may be affected by a change in tax rates.

¹⁴Welfarist optimal schedules are sensitive to the absolute levels of elasticities because low elas-

relative levels of elasticities across income levels. The pattern of elasticities by incomes is also poorly known empirically. Thus, for numerical simulations, the simplest is to assume that elasticities are constant along the income distribution. There is a stronger consensus in the empirical literature suggesting that income effects are likely to be small. Therefore, because of the theoretical simplicity of the no income effect case and its probable empirical relevance, I decided to present numerical simulations only for that case. Next Section considers the case with income effects from a theoretical point of view. Using those results to implement simulations in the presence of income effects would of course be possible.

With the assumption of constant elasticities, the empirical implementation of formula (2.12) is relatively easy. I assume that the government wants to raise the same amount of taxes per capita as it actually raises with present taxes, which is roughly one third of total wages or \$15,000 on average per household. In order to get rid of the endogeneity of the income distribution term in (2.12), I come back to the original Mirrlees formulation. I use utility functions with constant elasticity and the empirical wage income distribution to calibrate the exogenous skill distribution of the Mirrlees model. The details of this methodology are presented in appendix. Figure 4 displays the pattern of optimal marginal rates by income levels assuming elasticities are equal to 0.25 and using the 1992 wage income distribution to calibrate the skill distribution. The pattern of marginal rates is highly regressive, tax rates start at 100% at the bottom, then decline and reach a minimum of 6% at income level \$80,000 and then increase slightly up to an asymptotic value of 9%. I also report on Figure 4 the smallest flat tax rate (horizontal line on the Figure) which raises the same amount of taxes per capita. This flat tax level is equal to 35%. As evidenced on Figure 4, average non-linear tax rates are much smaller than the flat tax rate but the non-linear schedule imposes a very heavy burden on the poorest households. It is efficient for the government to impose high rates at the bottom because (almost) everybody has to pay these rates but the distortion is borne only by the low income people who would not have earned much in any case because they have low earnings potential. However, imposing the efficient non-linear tax rate on the poor households would probably be infeasible unless there is a negative income

elasticities imply that high levels of redistribution are less costly.

tax component. That is, the government would have give a guaranteed income level to everybody and then would tax it away using the rates displayed on Figure 4 (as discussed above, with no income effects, the guaranteed income level does not affect the shape of the optimal tax schedule).

As discussed above, a change in government revenue needs E affects the optimal pattern of tax rates mainly through a change in the marginal excess burden of public funds p . In order to display the effects of a change in E on tax rates, I have also reported on Figure 4 optimal tax rates for higher government revenue requirements. The case $E = \$27,500$ is nearly the maximum amount the government can raise using a non-linear income tax. This case corresponds thus to the revenue maximizing tax schedule. The case $E = \$22,500$ is intermediate between the revenue maximizing tax rate and the case $E = \$15,000$ examined above. When E increase, the tax rates for high incomes increase relatively more than the low income tax rates (which were already high) and the U-shape pattern of tax rates becomes more apparent. It can also be shown that p is a L-shaped function of E . When E is modest, p is very close to zero and increases very slowly. At some point, p start increasing quickly and tends to infinity when E tends to the maximum level of taxes the government can raise (Laffer tax). Similarly to Harberger's triangle formula, the marginal excess burden p is a highly non-linear function of the total tax burden per capita E

As a result, when revenue requirements are low (which is the case for ancient or underdeveloped economies), the optimal tax is close to the quasi poll-tax solution discussed in introduction where the tax burden is borne mostly by low incomes. However, when revenue requirements increase (as has been the case in developed countries in the 20th century) the government has to 'broaden' the tax base and start taxing high incomes at significant rates.

In Table 1, I compare the tax rates and total deadweight burden generated by the optimal non-linear tax schedule and the optimal flat tax for four different elasticity assumptions and with government revenue needs equal to \$15,000 per capita. First, as discussed earlier, changing elasticities affects very little optimal average tax rates (the pattern of optimal non-linear rates is also hardly affected by the levels of elasticities). Second and as shown in Figure 4, average rates are much lower for the non-linear case and even more so when average rates are computed using income weights instead of

population weights. Third, Deadweight burden levels are of course very sensitive to the elasticity levels; average deadweight burdens are between 3 to 5 times higher in the linear case than in the non-linear case. Last, note that the tax base is smaller in the linear case (especially when elasticities are high) because the flat tax distorts the economy more than the non-linear tax.

2.4 General Analysis

2.4.1 Deadweight Burden and Tax Reform

In order to analyze the general case with income effects, it is useful to consider individual deadweight burden as a function of z^* and u . Figure 1 illustrates this point very clearly. More precisely, Figure 1 shows that c^F and z^F depend only on u because (c^F, z^F) is the only point on the indifference curve u such that the tangent has slope one. Mathematically, (c^F, z^F) is determined uniquely by the equations $u = u(c^F, z^F)$ and $u_c(c^F, z^F) + u_z(c^F, z^F) = 0$. These two equations imply that (denoting $u_c^F = u_c(c^F, z^F)$),

$$u_c^F \left(\frac{\partial c^F}{\partial u} - \frac{\partial z^F}{\partial u} \right) = 1 \quad (2.14)$$

Once z^* and u are given, c^* is uniquely determined by the equation $u = u(c^*, z^*)$. Therefore, equation (2.5) shows that DWB can be considered as a function of z^* and u only. Noting $u_c^* = u_c(c^*, z^*)$ and $u_z^* = u_z(c^*, z^*)$, we also have $\partial c^*/\partial u|_{z^*} = 1/u_c^*$ and $\partial c^*/\partial z^*|_u = -u_z^*/u_c^* = 1 - T'(z^*)$. We can now compute the partial derivatives of DWB with respect to u and z^* . From (2.5) and (2.14), we obtain:

$$\frac{\partial DWB}{\partial z^*}|_u = \frac{\partial c^*}{\partial z^*}|_u - 1 = -T'(z^*) \quad (2.15)$$

$$\frac{\partial DWB}{\partial u}|_{z^*} = \frac{\partial z^F}{\partial u} - \frac{\partial c^F}{\partial u} + \frac{\partial c^*}{\partial u}|_{z^*} = \frac{1}{u_c^*} - \frac{1}{u_c^F} \quad (2.16)$$

Equations (2.15) and (2.16) are illustrated on Figures 5 and 6. Figure 5 shows that, for a given indifference curve u , increasing earnings z^* by dz^* reduces deadweight burden by the length of the starred segment which is approximately equal to τdz^*

where τ is the marginal tax rate at (c^*, z^*) . Figure 6 shows that, keeping z^* constant, and switching utility from u to $u + du$, changes deadweight burden by $dR - dA$. On Figure 6, dR (resp. dA) is the amount of income needed to increase utility by du starting from the bundle (c^*, z^*) (resp. (c^F, z^F)). Thus, dR and dA are such that $u_c^* dR = du$ and $u_c^F dA = du$ yielding the result in (2.16).¹⁵

Consider now a small tax reform (inducing a small change in the tax schedule) as displayed on Figure 7. Before the reform, a given individual was choosing the bundle (c^*, z^*) . The tax rate and the virtual income at the optimum (c^*, z^*) are equal to τ and R . The small tax reform induces a small change in earnings and consumption from z^* to $z^* + dz^*$ and c^* to $c^* + dc^*$. At this new optimum $(c^* + dc^*, z^* + dz^*)$, the marginal tax rate and the virtual income have been changed by $d\tau$ and dR . The change in consumption is such that $dc^* = dz^*(1 - \tau) + dR - z^*d\tau$. Because the condition $u_c^*(1 - \tau) + u_z^* = 0$ is satisfied at (c^*, z^*) , by the envelope theorem, the change in utility incurred by the individual from switching to (c^*, z^*) to $(c^* + dc^*, z^* + dz^*)$ is equal to:

$$du = u_c^* dc^* + u_z^* dz^* = u_c^* (dR - z^* d\tau) = u_c^* dC$$

As shown on Figure 7, $dC = dR - z^* d\tau$ is the change in after-tax income if there were no behavioral response: dC is the vertical distance between the two tax schedules at z^* .¹⁶ Therefore, using equations (2.15) and (2.16), the change in individual deadweight burden induced by the tax reform is,

$$dDWB = \frac{\partial DWB}{\partial u} du + \frac{\partial DWB}{\partial z^*} dz^* = \left[1 - \frac{u_c^*}{u_c^F}\right] dC - T'(z^*) dz^* \quad (2.17)$$

The term $T'(z^*) dz^*$ is the loss in tax receipts due to the behavioral response dz^* . The term dC is the loss in tax receipts due to the mechanical change in tax liabilities: $dC < 0$ means that tax liabilities increase. From formula (2.17), we see that the mechanical change in tax receipts dC changes DWB by dC times a factor $[1 - u_c^*/u_c^F]$.

¹⁵With no income effects, indifference curves are “parallel” and thus $dR = dA$, implying that DWB is independent of u in that case as can be seen from (2.16).

¹⁶This is only true on Figure 7 up to a second order term because the tax schedules are not necessarily linear. Nevertheless, this second order effect is one order of magnitude smaller than the first order effects I have described.

Noting $dTR = T'(z^*)dz^* - dC$ the total individual change in tax liabilities due to the tax reform, we have,

$$dDWB = -dTR - \frac{u_c^*}{u_c^F}dC \quad (2.18)$$

Summing equation (2.18) over all the individuals and using the notations of Section 3, we have,

$$dTDW = -dTTR - dU \quad (2.19)$$

where dU is the sum of the expression $(u_c^*/u_c^F)dC$ over the whole population. As in Section 3, at the optimum, for any small tax reform, changes in total deadweight burden and total tax receipts must be such that $dTDW = p dTTR$. Using (2.19), at the optimum, for any small tax reform, the following equation holds:

$$dU + (p + 1)dTTR = 0 \quad (2.20)$$

2.4.2 Links with Mirrlees Framework of Optimal Taxation

I discuss here the analogy with the Mirrlees problem of optimal taxation. Chapter 1 considers the same small tax reform as in Section 3 around the optimum tax schedule to derive conditions for optimality. As discussed above, the tax reform induces changes in tax receipts (denoted by $dTTR$) through behavioral responses in earnings and through mechanical changes in tax liability. The mechanical change in tax liabilities represents the change in tax receipts if there were no behavioral responses: this is the expression $dC = dR - z^*d\tau$ of Section 4.1 valid for any small tax reform. By the envelope theorem, the behavioral responses entail no first order effect on individual utility. However, the mechanical changes lead to first order effects on utility. The government aggregates these utility changes using welfare weights for each income level. I note dU' the resulting total change in welfare¹⁷ and p' the marginal value of tax receipts for the government. At the optimum tax schedule, $dTTR$ and dU

¹⁷More precisely, dU' is the sum of $\hat{g}dC$ over all the population where \hat{g} represents the marginal welfare weight.

must be such that (otherwise the tax reform or minus the tax reform would improve welfare),

$$dU' + p'dTTR = 0 \quad (2.21)$$

Equations (2.20) and (2.21) are obviously equivalent with $dU/(p+1) = dU'/p'$. Therefore, the tax schedule minimizing deadweight burden can be interpreted as a Mirrleesian optimal tax schedule with particular welfare weights. Given the expression for dU in (2.18), the welfare marginal weights $g(z^*)$ (in terms of the marginal value of tax receipts) for an individual with income z^* are given by,

$$g(z^*) = \left(\frac{1}{p+1} \right) E \left[\frac{u_c(c^*, z^*)}{u_c(c^F, z^F)} \middle| z^* \right] \quad (2.22)$$

where the operator $E[.|z^*]$ means expectation over all the individuals with income z^* . $g(z^*)$ is defined such that the government is indifferent (at the optimum) between giving an additional dollar to taxpayers with income z^* and having $g(z^*)$ more dollars of public funds. I show in appendix how this result can be obtained in the Mirrlees (1971) framework using his original maximization method. In the Mirrlees model, there is a single dimensional source of heterogeneity and so there is a single individual at each income level. Using the operator $E[.|z^*]$ is therefore not necessary in that case. Using the results of Chapter 1, the formula for tax rates minimizing deadweight burden can be written as:

$$\frac{T'(\bar{z})}{1 - T'(\bar{z})} = \frac{1}{\zeta_{(\bar{z})}^c} \left(\frac{1 - H(\bar{z})}{\bar{z}h(\bar{z})} \right) \int_{\bar{z}}^{\infty} [1 - g(z)] \exp \left[\int_{\bar{z}}^z - \frac{mpe_{(z')}}{\zeta_{(z')}^c} \frac{dz'}{z'} \right] \frac{h(z)}{1 - H(\bar{z})} dz \quad (2.23)$$

where $g(z)$ is given by equation (2.22). If the government has no redistributive tastes, it should be indifferent between small redistribution exchanges between any two individuals and thus the function $g(z)$ should be constant in z . However, when there are income effects, at the optimum, $g(z)$ is not constant in income.

Assume that leisure is a non inferior good then $mpe \leq 0$ (i.e., increasing non wage income reduces earnings) and that for a given tax rate and non wage income, the bundle (c^*, z^*) is optimal. Increasing non wage income shifts the budget up

and reduces optimal earnings. Thus the indifference curve crosses the new budget line at $(z^*, c^* + dR)$ from below as depicted on Figure 8. Therefore, with income effects, the vertical distance between two indifference curves increases as earnings increase (this result is easy to prove formally using calculus). The marginal utility of consumption at each income level u_c is inversely related to the vertical distance between the two indifference curves u and $u + du$ (see the comments on Figure 6 above). Therefore, along a given indifference curve, the higher earnings, the lower the marginal utility of consumption. With positive marginal rates, earnings are distorted downward ($z^* < z^F$) and thus the marginal utility of consumption is higher at the second best bundle than at the equivalent first best bundle: $u_c^* \geq u_c^F$. This result would be exactly reversed if leisure were an inferior good.

The consequence of this feature is that the measure of *DWB* varies as the utility level varies producing a varying pattern for welfare weights given by (2.22). With income effects, the welfare weights are always as large as $1/(p + 1)$. For a given individual, the higher the marginal tax rate he faces, the higher the distance between the second best and the first best and thus the higher the corresponding welfare weight.¹⁸ Computing the exact weight $g(z^*)$ requires knowing the first best level of earnings z^F . This requirement is not necessary in the case where welfare weights are considered to be constant at the optimum. Therefore, implementing numerically formula (2.23) is harder than implementing the same optimal tax formula but with constant weights. In any case, this rigorous analysis has shown clearly that abstracting from redistributive motives using *DWB* entail no significant simplification compared to the Mirrlees case. Moreover, in the presence of income effects, the solution of this problem is complicated by the fact that welfare weights are not constant. In that later case, the Mirrlees optimal income tax schedule with constant welfare weights makes more intuitive sense because no redistributive motives should mean exactly that the government is indifferent between taking away a dollar from a low income taxpayer to give it to a richer taxpayer (or the reverse). I come back to this point in the next subsection. This suggests that using the deadweight burden concept in the presence of income effects is problematic. Note however that these weights, though

¹⁸The welfare weight is exactly $1/(p + 1)$ if and only if the individual faces a zero marginal tax rate.

not necessarily constant, are always positive, implying that the tax schedule is in the class of Pareto efficient tax schedules.

Before concluding, I examine the link between the present result that minimizing deadweight burden generates non constant weights and previous findings that consumer's surplus arguments are severely limited in the presence of income effects.

2.4.3 The limits of Consumer's Surplus concepts

As discussed in the introduction, it is known since a long time that consumer surplus' theory and the deadweight burden concept may not make full sense unless 'the marginal utility of income is constant'. (see Samuelson (1964) for an early criticism of Harberber's (1964) method of computing excess burden). The case of Hick's compensating variation studied by Chipman and Moore (1980) is closely related to the present analysis. As displayed on Figure 9, the compensating variation is defined as the amount necessary to make somebody who faces a given tax rate τ and virtual income R as well off as in an initial situation (yielding utility u_0 say). Compensating variation can thus be seen as a function of $(1 - \tau, R)$. The question asked by Chipman and Moore (1980) is whether this function is an adequate indicator of indirect utility. Their main result is that compensating variation is adequate only in the case of "parallel" preferences with respect to the commodity which is used as numéraire. This corresponds precisely to the model of the present paper where deadweight burden is equivalent to using constant welfare weights only when preferences are "parallel" with respect to consumption (i.e., no income effects). The analysis of Chipman and Moore (1980) is purely theoretical but the intuition of their result can be easily understood looking at Figure 9. Bundle B yields a higher utility level than bundle A but the cost to compensate the individual from B to u_0 is higher than the cost to compensate the individual from A to u_0 . This counter-example is possible precisely because preferences are not "parallel". There is a close link between Figures 6 and 9. The welfare weights are not constant in the presence of income effects precisely because, as displayed on Figure 6, the measure of deadweight burden varies with the utility level: non-parallel preferences imply that dR is not equal to dA .

Except in the case of parallel preferences, changing prices will affect the marginal

utility of income. Therefore, it is impossible, using the classical welfare analysis, to assign equal marginal weights to everybody in such a way that that these weights remain equal when prices (or equivalently, taxes) change.¹⁹ The easiest way to assign the same weight to everybody is to force these weights to be the same for everybody at the optimum (i.e., impose a constant $g(z)$ function in formula (2.23)). This has the nice property that, at the optimum point, the government is indifferent between making some small transfer between any two individuals. One could use a utilitarian framework (using for example the Mirrlees (1971) model) where the government maximizes a weighted sum of individual utilities to generate this outcome. More precisely, the government would maximize $W = \sum_i \lambda_i U^i$ where U^i is the utility of individual i and λ_i is the weight associated to individual i . In order to generate constant marginal welfare weights (constant function $g(z)$), the λ_i must be chosen such that at the optimum, the $\lambda_i U_c^i$, which represent the social marginal utilities of consumption, are the same for everybody. Of course, the resulting weights λ_i are endogenous and if the government changes the tax schedule, the resulting social marginal utilities of consumption may not be constant any longer.

Consequently, welfare economics analysis has preferred to work with exogenous social welfare functions. But the drawback of that approach is that the problem of optimal taxation with no redistributive tastes cannot be defined in a completely satisfying way. More generally, welfare economics analysis with redistributive concerns has been considered as problematic because it involves interpersonal comparisons of cardinal utilities.²⁰ Thus it seems preferable to us to take as primitive the social marginal utility of consumption (which would be constant in the no redistributive concerns case) and work back the weights λ_i such that, at the optimum, the marginal utilities of consumption $\lambda_i U_c^i$ match the primitive social marginal utility of consumption chosen by the government. This method is completely independent of the cardinalization chosen to represent individual utilities U^i and has the advantage of displaying the redistributive concerns of the government in a completely explicit way. Moreover, as sketched above, this method can be made compatible with some direct social welfare function as in the traditional welfare analysis.

¹⁹Auerbach (1985) (pp. 82-83) develops this point in more detail.

²⁰See Stiglitz (1987), Section 7, for a more detailed exposition of this point.

2.5 Conclusion

This paper has derived the non-linear income tax schedule minimizing deadweight burden subject to a revenue requirement for the government. The features of this problem have been shown to be equivalent to the Mirrlees (1971) optimal income tax problem. The tax schedule minimizing deadweight burden is a Mirrleesian optimal tax schedule in which the government applies particular welfare weights at each income level. In the case of no income effects, these marginal welfare weights are the same for everybody: the tax schedule minimizing deadweight burden is then simply the Mirrlees optimal income tax with no redistributive concerns. The case with no income effects is exactly equivalent to the classical problem of nonlinear Ramsey pricing of the Industrial Organization literature. In the general case with income effects, these marginal welfare weights are higher for people facing higher marginal rates. This last property has been interpreted as a deficiency of the deadweight burden concept in the case of income effects. In any case, this paper shows that for the income tax problem, minimizing deadweight burden is not simpler than considering welfarist social criteria. In other words, adding redistributive concerns is straightforward once the pure efficiency analysis has been done. In particular, the paper has shown that the usual Harberger's triangle formula fails to capture key elements of the income tax problem. Even with no redistributive concerns, the shape of the income distribution has a strong impact on the pattern of marginal rates.

Numerical simulations have shown that there would be a significant loss of efficiency for the government to restrict itself to using a flat income tax rate in the case of constant elasticities. Given the shape of the income distribution, it is far more efficient to have a U-shape pattern of marginal tax rates with very high rates at the bottom and some progressivity at the top. However, the optimal efficient tax schedule is strongly regressive and could certainly not be implemented without a negative income tax (universal guaranteed income level) component.

Appendix

Analysis in the Mirrlees' Framework

I show in this appendix how the theoretical results of the paper can be obtained in the exact model of Mirrlees (1971) using his maximization methodology. In the Mirrlees' model, all individuals have identical preferences. The utility function depends on composite consumption c and labor l and is noted $u(c, l)$. The individuals differ only in their skill level (denoted by n) which measures their marginal productivity. If an individual with skill n supplies labor or "effort" l , he earns $z = nl$. The distribution of skills is written $F(n)$, with density $f(n)$ and support in $[0, \infty)$. The consumption choice of an individual with skill n is denoted by (c_n, l_n) and I write $z_n = nl_n$ for its earnings and u_n for its utility level $u(c_n, l_n)$. The government does not observe n or l_n but only earnings z_n . Thus it is restricted to setting taxes as a function only of earnings: $c = z - T(z)$. The government minimizes the sum of individual deadweight burdens:

$$TDW = \int_0^{\infty} DWB(n)f(n)dn \quad (2.24)$$

where $DWB(n)$ is the deadweight burden for an individual with skill n . In the Mirrlees (1971) model, the government maximizes a social welfare function:

$$W = \int_0^{\infty} G(u_n)f(n)dn \quad (2.25)$$

where G is an increasing and concave transformation of utility.

The government minimizes TDW (or maximizes W in Mirrlees (1971)) subject to a resource constraint and an incentive compatibility constraint. The resource constraint states that aggregate consumption must be less than aggregate production minus government expenditures, E :

$$\int_0^{\infty} c_n f(n)dn \leq \int_0^{\infty} z_n f(n)dn - E \quad (2.26)$$

The incentive compatibility constraint states that the selected labor supply l_n maximizes utility, given the tax function, $u(nl - T(nl), l)$, implying the first order condition $n(1 - T'(z_n))u_c + u_l = 0$. This first order condition leads to:

$$\frac{du}{dn} = -\frac{lu_l}{n} \quad (2.27)$$

Following Mirrlees (1971), in the maximization program of the government, u_n is regarded as the state variable, l_n as the control variable while c_n is determined implicitly as a function of u_n and l_n from the equation $u_n = u(c_n, l_n)$. I note $c = c(u, l)$ the implicit function defining c . Note that $\partial c/\partial u = 1/u_c$ and $\partial c/\partial l = -u_l/u_c$. Using equation (2.5), individual deadweight burden can also be expressed as a function of u_n and l_n :

$$DWB(n) = nl^F(u_n) - c(l^F(u_n), u_n) - nl_n + c(u_n, l_n) \quad (2.28)$$

where $l^F(u)$ is the first best level of labor supply on the indifference curve u . Similarly to equations (2.15) and (2.16), the derivatives of $DWB(n)$ can be written as:

$$\frac{\partial DWB(n)}{\partial l_n} = -n - \frac{u_l}{u_c} \quad (2.29)$$

$$\frac{\partial DWB(n)}{\partial u_n} = \frac{1}{u_c} - \frac{1}{u_c^F} \quad (2.30)$$

The government minimizes (2.24) by choosing l_n and u_n subject to equations (2.26) and (2.27). Forming a Hamiltonian for this expression, we have:

$$H = [-DWB(n) + p(nl_n - c_n)]f(n) + \phi(n)\frac{\psi(l_n, u_n)}{n} \quad (2.31)$$

where p and $\phi(n)$ are multipliers and $\psi(l, u) = -lu_l$. The Mirrlees (1971) Hamiltonian is the same as (2.31) with $G(u_n)$ replacing $-DWB(n)$. The first order conditions for the maximization of the Hamiltonian are given by $d\phi/dn = -\partial H/\partial u$ and $\partial H/\partial l = 0$. Using equations (2.29) and (2.30), we obtain:

$$-\frac{d\phi}{dn} = \left[\frac{1}{u_c^F} - \frac{p+1}{u_c} \right] f(n) + \phi(n) \frac{\psi_u}{n} \quad (2.32)$$

$$(p+1) \left[n + \frac{u_l}{u_c} \right] f(n) + \phi(n) \frac{\psi_l}{n} = 0 \quad (2.33)$$

In the Mirrlees (1971) case, with $G(u)$ replacing DWB in the Hamiltonian and noting

p' (instead of p) the budget constraint multiplier, the first order conditions (2.32) and (2.33) would be written as:

$$-\frac{d\phi}{dn} = [G'(u_n) - \frac{p'}{u_c}]f(n) + \phi(n)\frac{\psi_u}{n} \quad (2.34)$$

$$p'[n + \frac{u_l}{u_c}]f(n) + \phi(n)\frac{\psi_l}{n} = 0 \quad (2.35)$$

These two sets of equations are equivalent once we set $p' = p + 1$ and $G'(u) = 1/u_c^F$. The marginal weights that the government sets at each income level are given by $G'(u)u_c/p'$ in the Mirrlees case. Here, the corresponding weights are thus given by $(u_c/u_c^F)/(p + 1)$. The first order conditions (2.32) and (2.33) can be used to express optimal tax rates as a function of elasticities and the skill distribution exactly as done in Chapter 1. The only difference is that the welfare weights $G'(u)u_c/p'$ are replaced by $(u_c/u_c^F)/(p + 1)$. The final formula in terms in elasticities is equation (2.23) in the text.

Numerical Simulations

Numerical simulations are performed in the simple case with no income effects. Diamond (1998) specialized the Mirrlees model to that case with no income effects. He derived a formula equivalent to (2.12) but expressed in function of the exogenous skill distribution instead of the income distribution (equation (10) in Diamond (1998)). With constant welfare weights, this equation can be written as:

$$\frac{T'(z_n)}{1 - T'(z_n)} = \left(1 + \frac{1}{\zeta(z_n)}\right) \left(\frac{1 - F(n)}{nf(n)}\right) \left(\frac{p}{p + 1}\right) \quad (2.36)$$

where $F(n)$ is the skill distribution and z_n is earnings for an individual with skill n (see appendix for more details on the notation). This alternative formula is useful because it expresses the tax rates in terms of the exogenous skill distribution instead of the endogenous income distribution.

In the numerical simulations, I use utility functions with no income effects and constant elasticity (i.e., utility functions of the form $u = c - l^{k+1}/(k + 1)$ where the elasticity ζ is simply $1/k$). As in Chapter 1, the skill distribution $F(n)$ is then calibrated so that, given the actual tax regime, the resulting income distribution

matches the empirical wage income distribution of year 1992. Then, formula (2.36) is used to find p such that total tax receipts per capita (assuming $T(0) = -\$0$) be equal to the actual level of taxes E raised per capita on wage income. For year 1992, $E = \$15,000$. Once p is known, formula (2.36) gives the optimal tax rate for individual with skill n . Earnings are equal to $z_n = nl_n = n^{1+k}(1 - T'_n)^k$. Using this formula, the final curve $T'(z)$ can finally be computed and is reported on Figure 5.

The utility functions chosen satisfy obviously the Single Crossing property. From Mirrlees (1971), we know that equation (2.36) characterizes the optimum if the resulting earnings pattern z_n is increasing in n . In all the simulations, it has been checked that this is indeed the case.

Given the calibrated skill distribution, the optimal linear tax rate is simply computed as the smallest rate raising E per capita. Aggregate deadweight burdens are computed in a straightforward way once the optimal tax schedules are known.

Figure 1: Deadweight Burden Concept

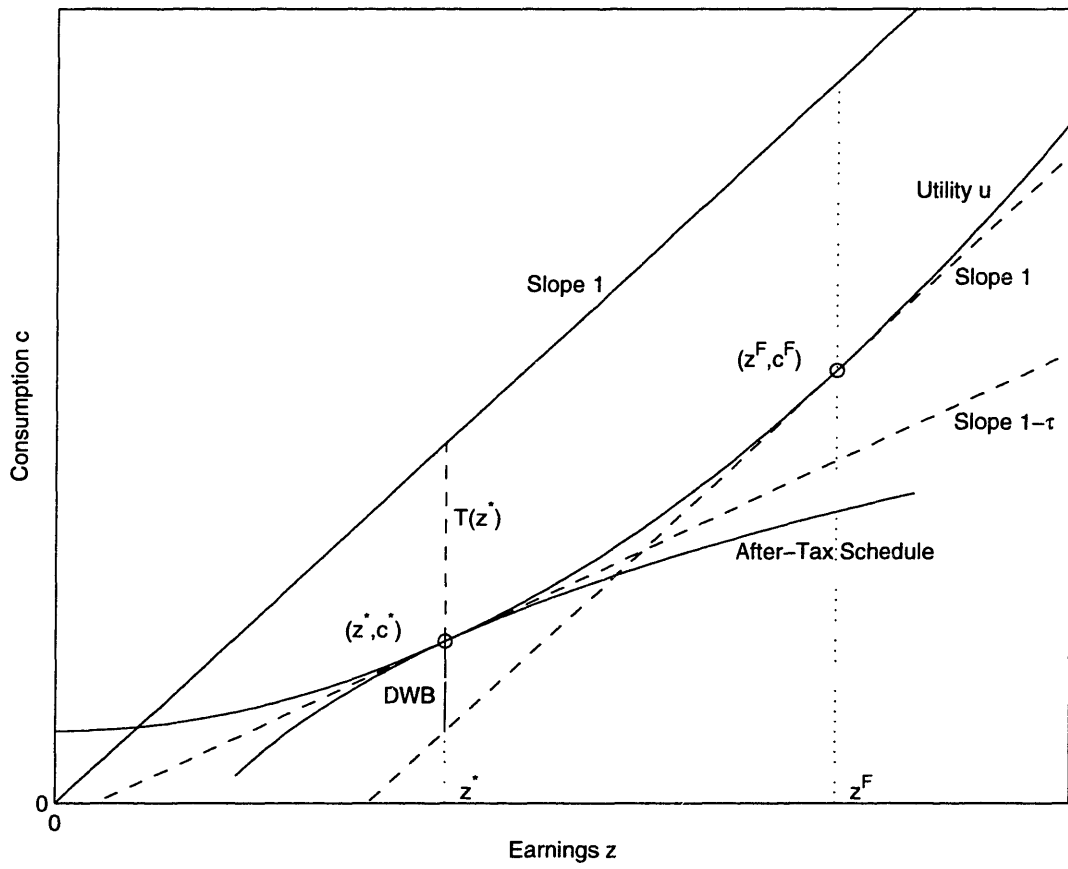


Figure 2: Deadweight Burden with No Income Effects

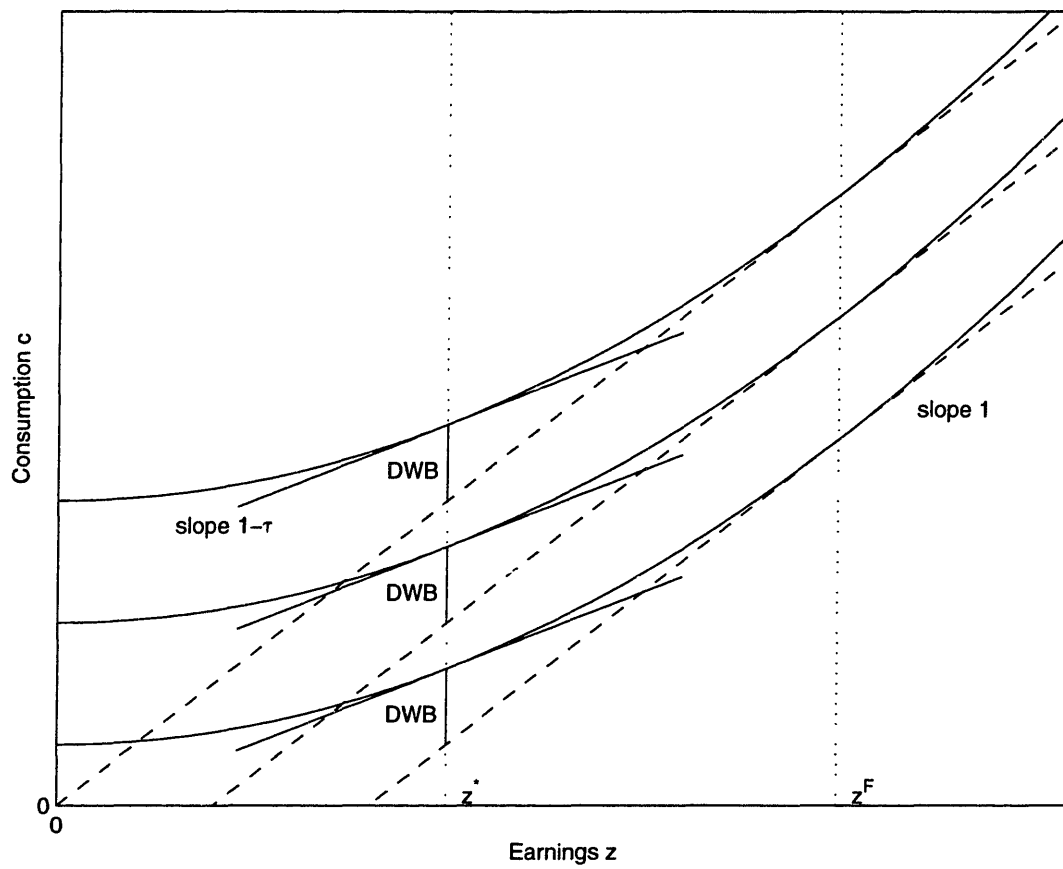


Figure 3: Ratio $(1-H)/(zh)$ for wages, years 1988, 1990 and 1992

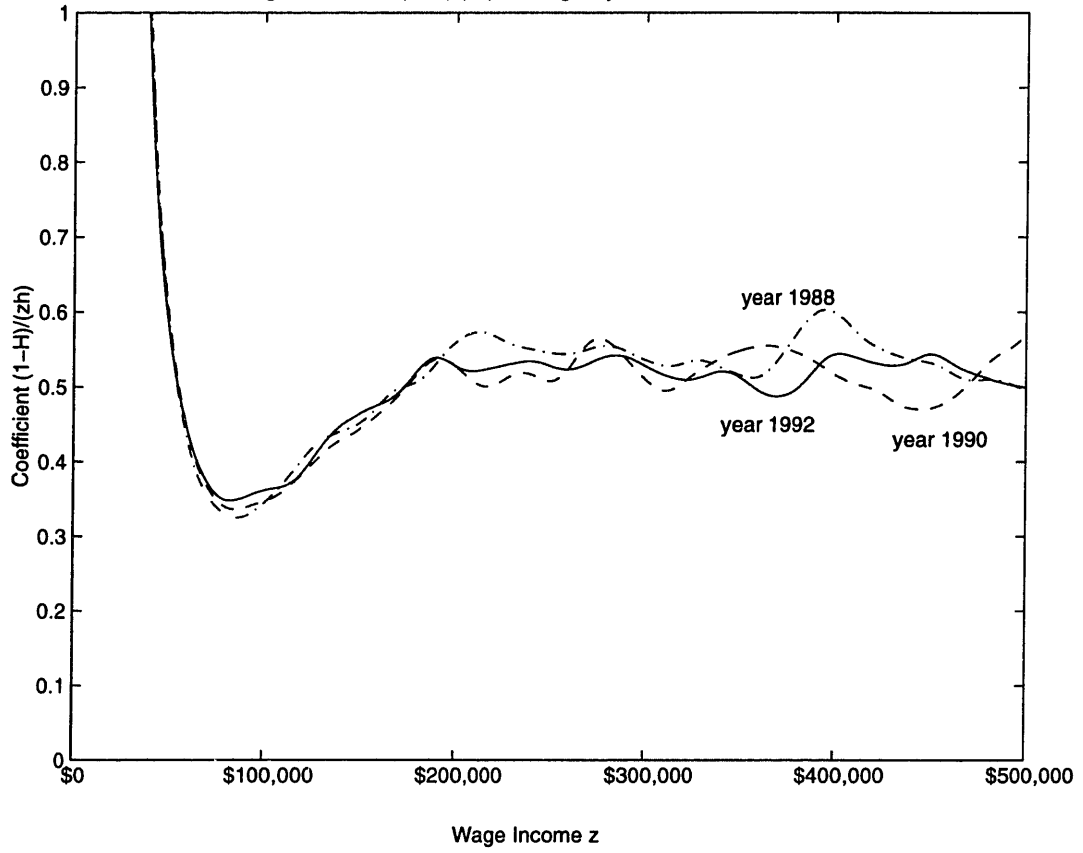


Figure 4: Optimal Tax Rates, Elasticity=0.25

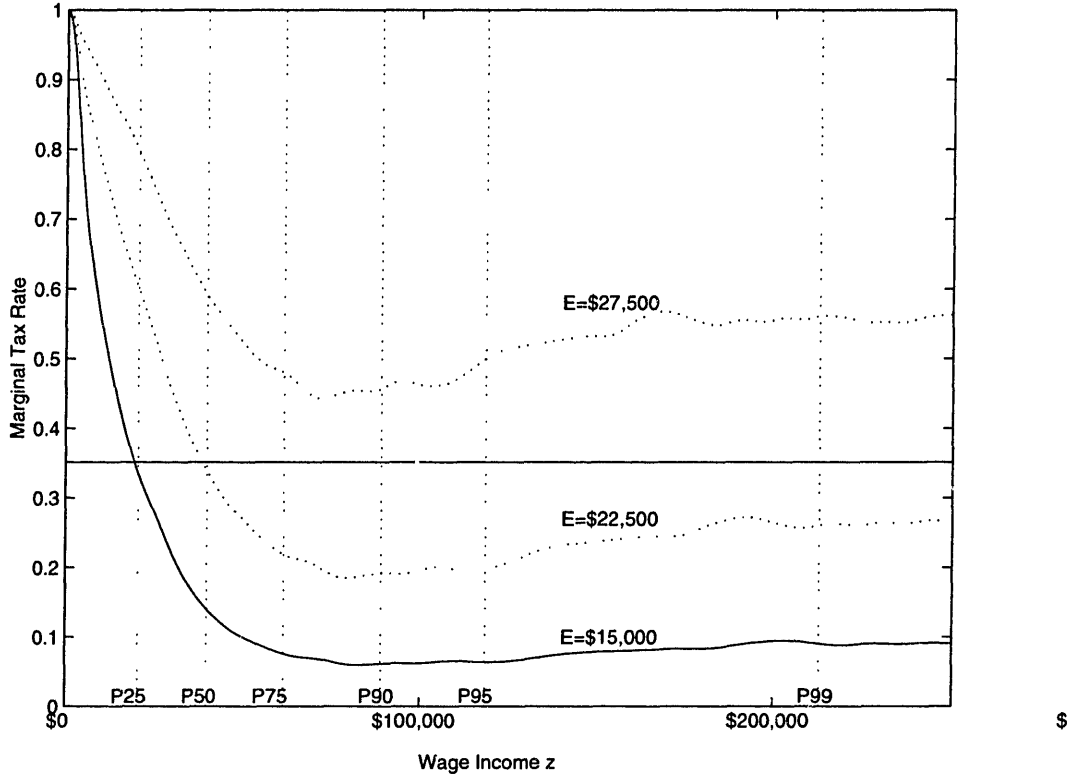


Figure 5: Derivative of Deadweight Burden with respect to earnings z

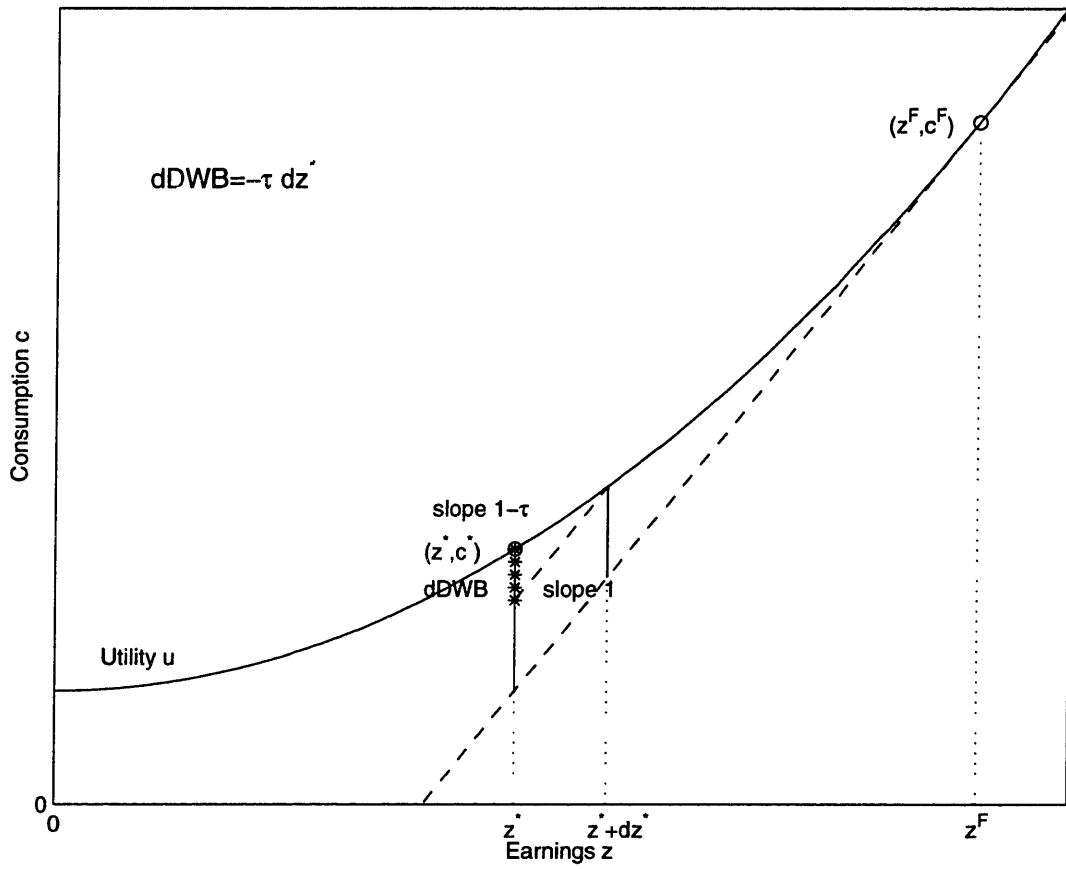


Figure 6: Derivative of Deadweight Burden with respect to utility u

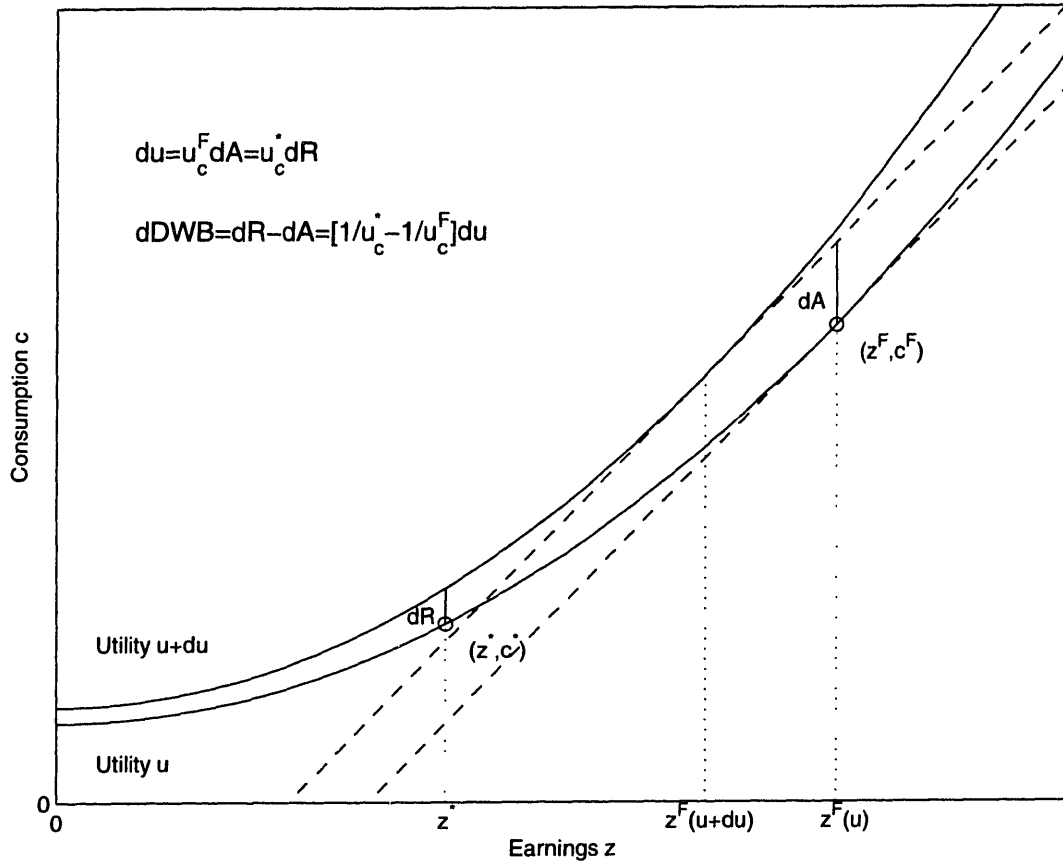


Figure 7: Response to a Small Tax Reform

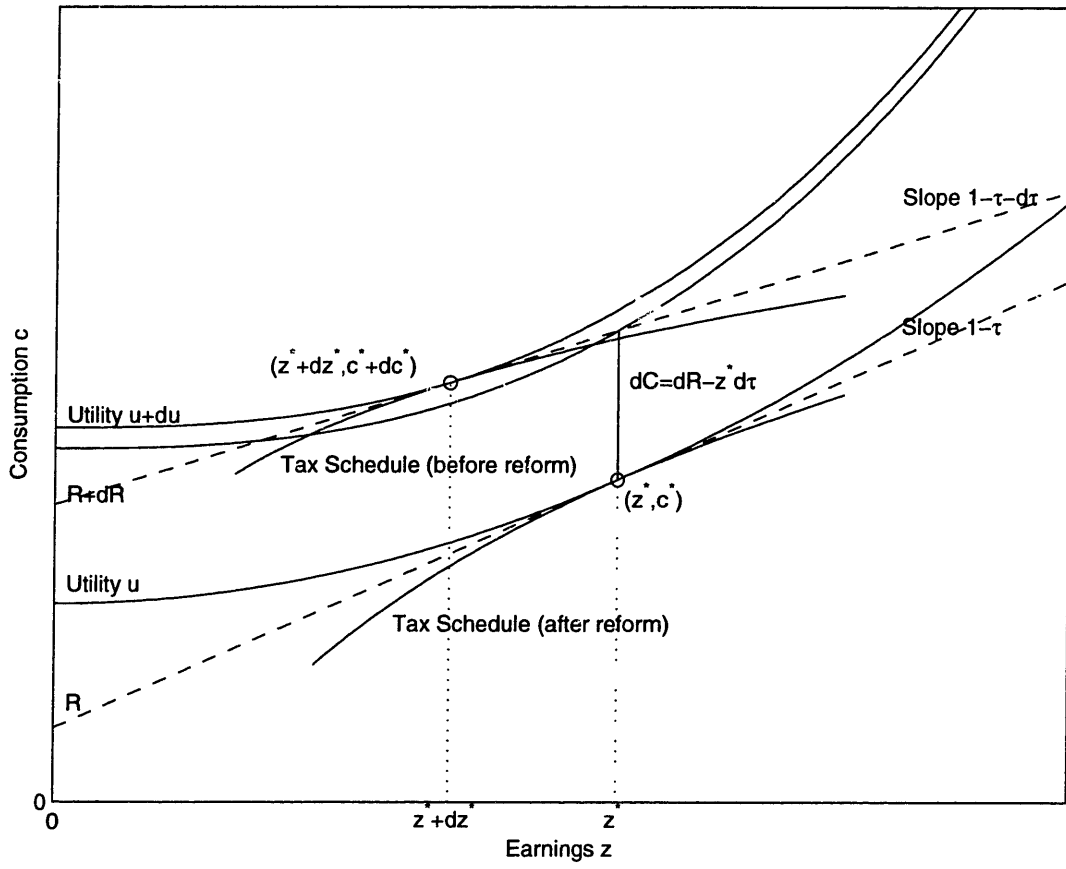


Figure 8: Income Effects and Indifference curves

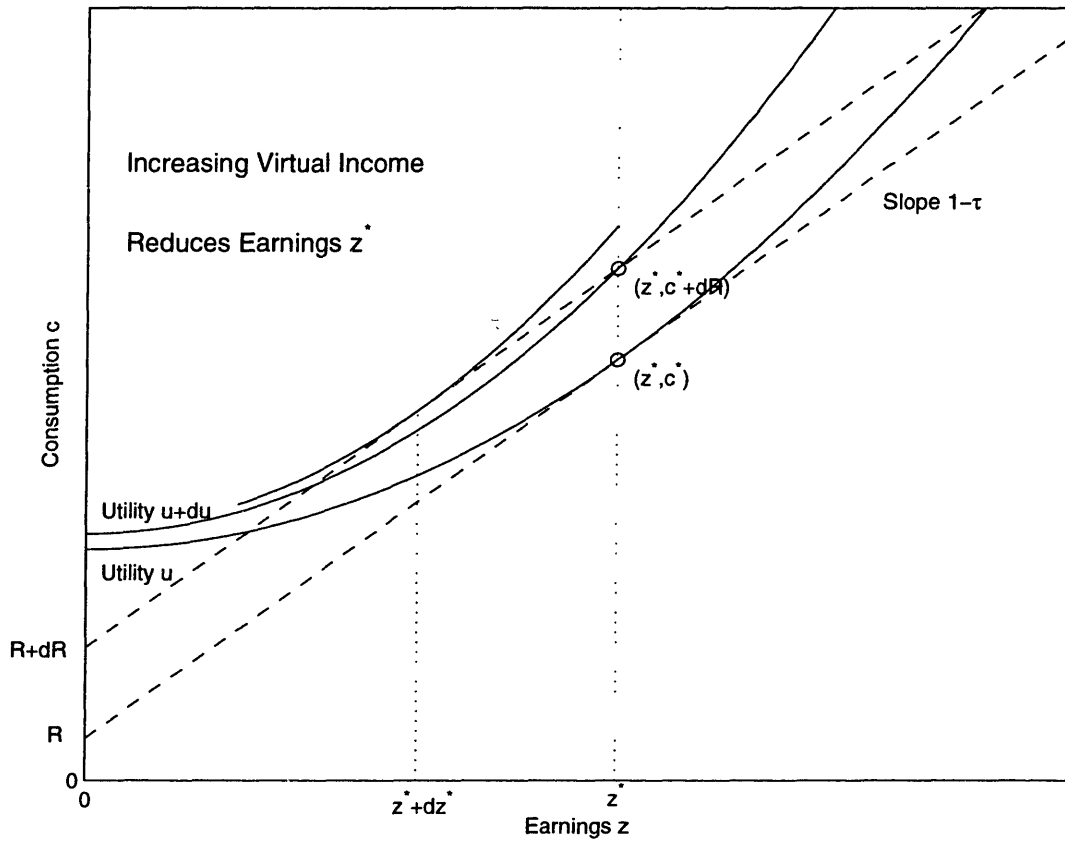


Table I: Summary Statistics of Tax Schedules Simulations

	Elasticity Parameter							
	0.25		0.5		0.75		1	
	Non-linear (1)	Linear (2)	Non-linear (3)	Linear (4)	Non-linear (5)	Linear (6)	Non-linear (7)	Linear (8)
Optimal Average Rate (income weights)	12.4	35.2	12.8	35.7	12.8	37.2	12.4	40.0
Optimal Average Rate (population weights)	25.0	35.2	25.9	35.7	26.5	37.2	26.3	40.0
Asymptotic Rate in non-linear case	9.0		9.7		10.0		9.9	
Average Deadweight burden (in dollars)	268	906	517	1,910	796	3,190	1,040	4,980
Average Deadweight burden (over total tax receipts E)	1.8	6.0	3.5	12.7	3.3	21.3	6.9	33.2
Marginal cost of Public Funds (p)	0.050		0.114		0.191		0.267	
Ratio Tax Bases (linear over non-linear)	0.932		0.867		0.792		0.697	

Notes: The figures presented in this table correspond to the simulations described in the text. The simulations are based on the wage income distribution of year 1992 for married households. The required amount of public funds is $E=\$15,000$. Deadweight burden figures are expressed in year 1992 dollars.

Chapter 3

The Effect of Marginal Tax Rates on Income: A Panel Study of ‘Bracket Creep’

3.1 Introduction

The response of taxpayers to changes in marginal rates has long been of interest to economists. The magnitude of this response is of critical importance in the formulation of tax policy and the determination of the size of the government and welfare programs. However the empirical literature has failed to generate any consensus on the magnitude of the elasticities of income with respect to marginal tax rates: estimates range from no effect to extremely large effects.

The labor supply literature focuses mostly on the elasticity of hours of work with respect to marginal tax rates and finds in general small responses to taxation. This literature suffers from two major drawbacks. First, hours of work might not be the only dimension of the total behavioral response to taxation, which is the relevant variable for tax policy purposes. Second, the identification of elasticities in the labor supply literature rests in general on strong functional form assumptions. Estimates are therefore sensitive to these functional form assumptions.

Recent studies have looked directly at the sensitivity of overall income with respect to marginal rates using tax reforms to identify the parameters of interest. These stud-

ies have used the US tax reforms of 1981 and 1986 to estimate taxpayers' responses. They find very large responses to taxation. This recent literature also suffers from major problems. First, the tax reforms introduced many changes in the definition of taxable income besides tax rate changes and thus it is often problematic to compare reported income before and after the tax reform. Second, these studies compare high income taxpayers (who experienced large tax rate cuts) to low and middle income taxpayers (who experienced almost no tax rate changes). Therefore, this methodology amounts to attributing the widening in inequalities to the tax reform. Third, this literature is not able, as opposed to most labor supply studies, to tell apart income and substitution effects. The knowledge of the size of each of these effects is important for tax policy.

These objections suggest that a research design to estimate behavioral responses to marginal tax rates should meet two conditions. First, the tax change should affect only marginal tax rates without introducing many changes in tax rules. Second, the tax change should affect differently groups of taxpayers that are comparable (i.e., whose incomes and other economic characteristics are close). The 'bracket creep' in the US income tax of the early eighties is a tax change meeting these two conditions.

From 1979 to 1981, inflation was high (around 10%) but the tax schedule was fixed in nominal terms. Because the income tax was highly progressive-there were about 15 tax brackets with rates increasing from 0 to 70%-inflation had a strong real impact.¹ The kink points of the tax schedule, fixed in nominal terms, shifted down in real terms. Therefore, a taxpayer near the top-end of a bracket was likely to *creep* to the next bracket even if his income did not change in real terms. The other taxpayers (far from the top-end of a bracket), however, were not as likely to experience an increase in marginal rates the following year. This characteristic of 'bracket creep' is exploited in this study to estimate the elasticities of income with respect to marginal rates. The spirit of the empirical strategy is to compare changes in income of taxpayers near the top-end of a bracket to changes in income of other

¹The effect of bracket creep on the US income tax was so strong that it increased substantially the average marginal rates and was the main cause of the 'tax revolt' of the late 1970s and early 1980s (see Steuerle (1991), Chapters 2 and 3, for a more detailed discussion). By comparison, the income tax cuts of 1981-84 were in fact just enough to bring total federal income tax receipts over GNP back to their 1977 level.

taxpayers.

This identification strategy has three advantages relative to the tax reform experiments of the eighties. First, I compare groups of taxpayers whose incomes are very close. Therefore, the estimates are likely to be robust to changes in the underlying distribution of income and in particular to underlying increases in inequality. Second, the ‘bracket creep’ phenomenon did not modify the definitions of reported income and thus incomes can be easily compared across years. Third, as a theoretical matter, I will show that the estimates obtained using ‘bracket creep’ are not a mix of income and substitution effects but rather pure compensated elasticities of income with respect to marginal tax rates. Three other important characteristics of the ‘bracket creep’ tax change should be mentioned. First, because I compare year to year changes, my study will capture only short term responses to tax changes which might be different from medium or long term responses. Second, changes in tax rates due to ‘Bracket Creep’ were relatively small compared to the changes induced by the large tax reforms of the eighties and thus it is harder to obtain precise estimates. Last, because ‘bracket creep’ was not a legislated change, it might have been harder for taxpayers to understand the effect of this change on marginal tax rates. I come back to these important points in more detail in the concluding Section.

The paper proceeds as follows. Section 2 briefly reviews the empirical literature on behavioral responses to taxation. Section 3 presents in detail the effects of ‘bracket creep’ on the tax schedule. This study requires the precise location of taxpayers on the tax schedule and also requires following taxpayers over several years. Therefore I use a publicly available panel dataset of tax returns constructed by the Internal Revenue Service (IRS). The dataset, summary statistics and raw differences-in-differences results are presented. Section 4 introduces the regression framework and specification and Section 5 displays the regression results. Section 6 presents caveats, discusses policy implications and concludes.

3.2 Literature

The basic approach of the traditional labor supply literature was to posit a linear budget constraint and regress hours of work on (after-tax) wage rates and non-wage

income. This literature has in general found very small elasticities (both compensated and uncompensated) of labor supply with respect to wages rates (or equivalently to marginal tax rates) for prime age males. Pencavel (1986) is an extensive survey of these studies. Estimates for the uncompensated elasticity are usually slightly negative (around -0.1). The compensated elasticity estimates are in general slightly higher but usually below 0.2.

Hausman (1981) applied a new methodology taking full account of the non-linearity of the budget set due to the progressive structure of the US income tax and challenged the prevailing wisdom that taxes had almost no incentive effect on labor supply. This non-linear budget set methodology has been used in many papers to estimate labor supply elasticities (these studies are surveyed in Hausman (1985)). These studies tend to find small uncompensated elasticities but high income effects leading to substantial compensated elasticities (often around 0.5). Non-linear budget set estimates have been shown to be sensitive to small changes in specification (see MaCurdy *et al.* (1990) and Triest (1990)).

Both the traditional labor supply literature and non-linear budget set studies suggest consistently that the elasticity is larger for secondary earners (married women): the elasticities found are often between 0.5 and 1 (e.g. Hausman (1985), Mroz (1987), Heckman (1993)).

The labor supply literature has been criticized along various lines. First, the estimates are dependent on the functional form chosen for the statistical inference. In other words, the identification of the key parameters comes from strong structural form assumptions. Note however that, because of these strong structural assumptions, labor supply studies can in general estimate both income and substitution effects. Second, hours of work may not be the only dimension of “effort”: individuals can vary their labor supply in the short run not only by changing hours but also by changing the intensity of work. In the long run taxpayers can also change the types of job they choose (see Feldstein (1995) for a more detailed discussion of this point). What matters for tax policy is the total response of reported income with respect to tax rates. Therefore, labor supply estimates may be substantially lower than the relevant total income elasticity.

Looking directly at the income response of taxpayers to tax reforms seems to be

a more promising approach to solve these two problems. First, tax reforms provide an *exogenous* time variation in marginal tax rates so that weaker functional forms assumptions can be used to identify the parameters of interest. Second, it is possible to study directly the total income response without need to focus only on hours of works. Previous research connecting the changes in reported income to changes in marginal tax rates include Lindsey (1987), Navratil (1995), Feldstein (1995) and Auten and Carroll (1997). The first two studies used the tax cuts of the Economic Recovery Tax Act (ERTA) of 1981 and the last two used the Tax Reform Act (TRA) of 1986 to identify the elasticities. All four studies used Internal Revenue Service (IRS) datasets of tax returns. The last three studies used a panel of tax returns whereas Lindsey (1987) had to use a repeated cross section because the panel was not yet available at the time he made his study.

Lindsey (1987) ranked the individual taxpayers by adjusted gross income before the ERTA and after the ERTA. His key assumption was that the successive fractiles corresponded to the same individuals in both years. He then related the change in average income for successive fractiles to predicted changes in their marginal net-of-tax rates (i.e., one minus the marginal rate). Lindsey's analysis implied very large elasticities: between 1 and 3, his preferred estimate being equal to 1.6.

Navratil (1995) used instead a panel of tax returns and was therefore able to test the critical assumption of Lindsey. He found that income mobility is quite important and argued convincingly that Lindsey's assumption leads to dramatically upward biased results. Navratil compared years before the ERTA and year 1983 after the ERTA. He derived his elasticity estimates by regressing the log change in income on the predicted log change in net-of-tax rates.² It is very important to note that this methodology does not lead to real elasticity estimates because this is a reduced form regression. To get estimates of the elasticity of income with respect to marginal tax rates, Navratil should have regressed the log change in income on the real log change in rates using the predicted log change in rates as an instrument (I discuss this point again in Section 4).³ Navratil finds overall elasticities of about 0.8 for taxable income

²The predicted log change in net-of-tax rates is equal to $\log(1 - t_1/1 - t_0)$ where t_0 is the marginal rate before tax reform and t_1 is the post-reform marginal rate at the *before* tax reform taxable income level, adjusted for inflation.

³The study of Lindsey suffers from the same reduced form problem.

and about 1 for wages and salaries. These estimates are smaller than Lindsey's but still very high compared to the labor supply literature.

Feldstein (1995) uses a similar methodology with the TRA of 1986 and the same panel of tax returns.⁴ Feldstein divided his sample into three income groups: medium, high and highest. These groups experienced different marginal rates cuts. The tax cuts were smaller for medium income earners than for high income earners. The cuts for highest income earners were even larger than for high income earners. Feldstein then computes the change in taxable income between year 1985 and 1988 for each group and derives a differences-in-differences elasticity estimate by comparing changes across the different groups. As high income earners experienced a larger increase in revenue than low income earners, Feldstein obtains high elasticity estimates (ranging from 1 to 3). Some of Feldstein's results are based on very small samples and therefore the estimates are probably not precise (see Slemrod (1996) for a discussion of this point). These estimates are again reduced form estimates. The analysis is also complicated by the fact that the TRA of 1986 introduced many changes in tax rules and therefore the definitions of adjusted gross income and taxable income were substantially modified. The large decrease in upper marginal rates may also have induced many wealthy taxpayers to shift corporate income which was taxed as corporate profits to S corporations and partnerships which are taxed as personal income tax (see Feldstein (1995) and Slemrod (1996) for a discussion of this point).

As pointed out by Navratil (1995) (Chapter 2), Feldstein's results depend critically on the assumption that the elasticities are the same for the three groups. Navratil computes elasticity estimates based on a simple pre-post reform comparison for each of the three groups. The three estimates are very different (though not statistically different because of the small size of the sample) but the three of them are substantially smaller than Feldstein's differences-in-differences estimate.

Auten and Carroll (1997) repeated the study of Feldstein but with a much larger panel dataset of tax returns available only to researchers at the US Treasury. They compute structural estimates using an instrumental variable method. They are also able to control for some non-tax factors such as age, state of residence and type of

⁴In fact, Feldstein's study preceded Navratil's study and thus was the first one to use the panel data of the IRS to estimate elasticities of taxable income with respect to marginal rates.

job. They obtain smaller estimates than Feldstein: their preferred estimate is equal to 0.66. It is however difficult to compare directly their results with Feldstein's because they present neither their first stage estimates nor the reduced form estimates that Feldstein reported.

The most important problem with the studies reviewed above is that the marginal rate cuts of the two Tax Reforms (ERTA and TRA) increased with income: wealthy taxpayers experienced larger marginal rates cuts than poorer taxpayers. Therefore, imputing the faster increase of high incomes compared to low incomes only to the tax reforms leads to upward biased estimates if increases in inequality are partly due to other factors than tax cuts. Economists have proposed many other explanations for increased income inequality: Murphy and Welch (1992) and Katz and Murphy (1992) found that the returns to human capital or education increased as a result of increased demand for skilled labor. Declining union membership (Freeman (1993)), increasing import competition (Bound and Johnson (1992)), increasing immigration (Topel (1994)) have also been proposed as potential explanations of the widening inequalities over the last 25 years in the US.

We have seen that the recent tax response studies have a decisive advantage compared to old structural labor supply studies because the identification problem is not solved artificially through strong functional form assumptions. However, this advantage has a cost: the tax reform studies are no longer able to tell apart substitution and income effects. These studies present a single elasticity estimate which is neither a pure compensated elasticity nor a pure uncompensated elasticity but a mix of both elasticities. In general, the studies using legislated tax reforms do not discuss this issue at all. It is important, though, to be able to tell apart each elasticity because optimal tax rates levels depend on the size of both elasticities (see Chapter 1).

My paper will try to address these issues. After describing in details the tax changes due to 'bracket creep', I argue in Section 3.1 why my estimates are free from the problems affecting the existing literature about the behavioral responses to taxation.

3.3 ‘Bracket Creep’, Data and Descriptive Statistics

3.3.1 The ‘Bracket Creep’ phenomenon

The analysis presented here uses the same panel of tax returns as Feldstein’s and Navratil’s studies but does not use a tax reform to carry out the estimation. The paper focuses instead on a very different kind of tax change. From 1979 to 1981, the tax schedule was not indexed even though inflation was on the order of 10% per year. Non-indexation changed the tax schedule because the income tax was highly progressive; this phenomenon was called ‘bracket creep’.

Figures 1 and 2 show the effect of inflation on the tax schedule and on marginal rates. After tax real income as a function of before tax real income is represented on Figure 1 for two consecutive years: the straight line represents the year 1 schedule and the dashed line the year 2 schedule. The kink points (i.e., the points where the marginal rate jumps) shift to the left because of inflation, but the slopes of the segments linking the kink points do not change. The marginal rates schedules are represented on Figure 2. If taxable income remains the same in *real* terms in year 2, then some taxpayers will face a higher rate: this is the “treatment” group. The other taxpayers will still face the same rate: this is the “control” group. These different groups are displayed on the figures. Formally, if the tax schedule is given by T in year 1, then in real terms, the tax schedule in year 2 is \hat{T} defined by:

$$\hat{T}(x) = T[x(1 + \pi)]/(1 + \pi)$$

where π is the inflation rate, and x is real income. Therefore,

$$\hat{T}'(x) = T'[x(1 + \pi)]$$

The tax changes induced by ‘Bracket creep’ have several advantages compared to the studies of Lindsey, Navratil and Feldstein reviewed above. First, there were almost no changes in the income tax code during the three years I focus on, therefore the only change is due to inflation. Comparisons across years are thus straightforward

compared to the tax reforms studies.

Second, and more importantly, kinks are regularly spaced along the whole income distribution. Therefore, control and treatment groups can be constructed over a large portion of the income distribution. Also noteworthy is the fact that controls and treatments *alternate* and thus for a given kink the treatment group and the two surrounding control groups are very similar in terms of income and very likely to share the same economic characteristics. Therefore the difference in changes in income between these groups can be confidently attributed to marginal rates effects. The estimates are thus likely to be robust to changes in the distribution of income and especially to changes in inequality.

Last, I will show in Section 4 that the elasticity estimates obtained using ‘bracket creep’ are in fact compensated elasticities of income with respect to marginal tax rates. Therefore, the usual deadweight burden approximations (which involve only the compensated elasticity) measuring the welfare costs of taxation could be easily computed. More generally, it is important for optimal income tax purposes to know the size of both compensated elasticity and income effects.⁵ The analysis of ‘bracket creep’ provides estimates of the first of these two key parameters.

However, the changes in marginal rates are not very large because there were many kink points at that time and the jumps in marginal rates were in general of 4-7% (see below). This is small compared to a decrease from 50% to 28% in marginal rates for the very high income earners following the TRA of 1986. However, Steuerle (1991) provides evidence that the ‘bracket creep’ of late 1970s and early 1980s was perceived as a major tax event. ‘Bracket creep’ triggered the strongest increase in marginal tax rates since World War II in just a few years. Federal income tax receipts over GNP increased very quickly from 1978 to 1981. According to Steuerle, this was the main cause of the ‘tax revolt’ and the tax cuts which took place in the 1980s. As 1980 was not the first experience of ‘bracket creep’ in the US (inflation was also high in the 1973-1975 period), it is very likely that ‘bracket creep’ was noticed and understood by most taxpayers.

⁵Chapter 1 shows that optimal income tax formulas can be expressed in terms of these two parameters and the shape of the income distribution.

3.3.2 Data

The IRS panel of tax returns which I use in this study covers the period 1979 to 1990. However, only the first three years are used for this project. This panel, known as the Continuous Work History File, contains most items on Form 1040, as well as numerous other items from the other forms and schedules. The IRS panel is constructed from all tax returns filed in a given year by selecting certain 4-digit endings of the social security number of the primary taxpayer listed on the form. Five such 4-digit endings were selected in 1979-1981, the three years used in this study. For each of these years, the panel contains about 46,000 observations. Due to budgetary limitations, only one 4-digit ending was chosen in 1982 and 1984 and two 4-digit endings were chosen in the other years. Thus Feldstein's and Navratil's studies were based on relatively small samples. After several deletions, Navratil used about 2,000 observations and Feldstein about 3,500.

Attrition in the panel can occur due to late filing or no filing (which can happen for example if the taxpayer does not owe any taxes and does not expect a refund from the IRS). Attrition may also result from a change in marital status if the name of the primary taxpayer listed on the return changes (see Christian and Frischmann (1989) for a more complete discussion of attrition in this panel).

In the US, there are different tax rate schedules for taxpayers filing as Singles, Married⁶ or Heads of household). As singles and married constitute about 90% of all tax returns, I will consider only single and married taxpayers. I compare year 1980 to year 1979 and year 1981 to year 1980. These two differences are stacked to obtain a dataset of about 80,000 observations. I then exclude taxpayers whose marital status changes from year 1 to year 2. It is unlikely that 'bracket creep' affected specifically marriage strategies and therefore discarding those observations should not bias the results. I also exclude taxpayers who do not use the regular tax schedule in year 1.⁷

⁶Married taxpayers can choose to fill either jointly or separately. The overwhelming majority of married taxpayers (more than 98%) chooses to fill jointly. Therefore, married taxpayers filing separately will be not be considered in my study.

⁷Most of these excluded taxpayers used the average income tax schedule which allowed taxpayers to replace their taxable income by an average of the last few years taxable income. This reduced the tax liability of taxpayers who had experienced a sharp rise in income. I also exclude taxpayers using the Maximum Tax Rate on Personal Service Income. The aim of the Maximum Tax Rate was to constrain the top rate on earned income to 50% (instead of 70%).

Real growth of GDP was small in 1980 and 1981: -0.5% in 1980 and 1.8% in 1981. The GDP deflator was 10.5% in 1980 and 9.5% in 1981. These figures are very close to the nominal growth of adjusted gross income for each year. The results I present are not sensitive to small changes in these parameters, which I call the “inflation parameters”. Most items reported on tax returns can be considered to grow roughly at the inflation rate. This is the case for adjusted gross income (AGI), wages and salaries, itemized deductions. Therefore I can express these items for year 2 in year 1 dollars just by dividing them by the inflation rate.

Taxable income is the key item to divide the sample into control and treatment groups. Taxable income is computed in two different ways depending on whether the taxpayer itemizes deductions or chooses the standard deduction. A taxpayer itemizes when the total of his itemized deductions is larger than the standard deduction. The standard deduction is fixed in nominal terms: 3,400 dollars for married taxpayers and 2,300 dollars for singles. If the taxpayer does not itemize, taxable income is simply equal to AGI minus personal exemptions.⁸ If the taxpayer itemizes, taxable income is equal to AGI minus personal exemptions minus itemized deductions plus the standard deduction.⁹

I write $taxinc_i$ for *nominal* taxable income in year i . $taxinc_i$ is simply taxable income reported on the tax form in year i . To assign a taxpayer to a treatment or control group, I compute predicted taxable income ($taxinc_p$) which is $taxinc_1$ expressed in year 2 dollars. If the marginal rate corresponding to $taxinc_p$ is above the one corresponding to $taxinc_1$,¹⁰ the observation is assigned to the treatment group of the corresponding kink. If the marginal rates for $taxinc_1$ and $taxinc_p$ are the same, the taxpayer is assigned to the control group. In order to compute the real change in taxable income, I also express $taxinc_2$ in terms of year 1 dollars (this is denoted by $taxinc_{2R}$). The details of the computations of $taxinc_p$ and $taxinc_{2R}$ are given in appendix. From now on, I denote by $T'_i = T'(taxinc_i)$ the effective marginal rate in year i and $T'_p = T'(taxinc_p)$ the predicted marginal rate in year 2 if *real* income does not change.

⁸Exemptions were fixed in nominal terms: 1,000 dollars for each person in the household.

⁹The definition of taxable income changed after the Tax Reform Act of 1986. The standard deduction is no longer included in taxable income and the zero tax bracket has disappeared.

¹⁰That is, by reporting the same real taxable income, the taxpayer would *creep* to the next bracket.

3.3.3 Descriptive Statistics

Figures 3 and 4 display the actual marginal rate schedules of year 1979 and the real effect of ‘bracket creep’ on tax rates for married and single taxpayers respectively. These figures are the empirical counterpart of Figure 2; the nominal location of kink points are reported (in thousands of 1979 dollars) on the horizontal axis, the marginal tax rates are reported on the vertical axis. The solid line represents the nominal schedule for year 1 while the dashed line represents the *real* schedule in year 2 (assuming a 10% inflation rate). Tables I and II show the summary statistics for each control and treatment group, for married and single filers. The groups are ordered by increasing taxable income in year 1. For each kink, the nominal level of taxable income at which the kink takes place and the jumps in marginal rates are presented in columns (2) and (3). Therefore, these Tables describe fully the tax schedule of years 1979 to 1981 for married taxpayers and single taxpayers. There were 15 kinks for married taxpayers and 16 kinks for singles.¹¹ I have constructed two control groups with incomes below the first treatment group (Control N and Control 0) in order to emphasize the mean reversion phenomenon for very low incomes. I have discarded the observations below Control N because taxpayers who report very low (or even negative) taxable income are often middle-high income earners which have faced a transitory sharp decline in taxable income. Slemrod (1992) discusses this point in detail. I indicate the number of observations for each group in column (4). The number of observations decreases quickly for the highest kink points because the panel does not overweight wealthy taxpayers.

Next, in column (5), the log ratios of predicted net-of-tax rates ($\log[(1 - T'_p)/(1 - T'_1)]$) are reported. The values are equal to zero for the controls because by definition, the marginal rate they face remains the same in year 2 if their *real* taxable income does not change. For treatments the values are negative: e.g. for the treatment corresponding to the kink 37/43 the value reported is $\log[(1 - 0.43)/(1 - 0.37)]$. This is the log change in net-of-taxes rates that a taxpayer in the corresponding treatment group would face if his *real* taxable income did not change from year 1 to year 2.

¹¹I have not reported statistics for the last jump in marginal rates from 68% to 70% because the size of this jump is small and there are very few observations around that last kink point. The last control group is composed of taxpayers below the treatment group for the kink 68/70.

Column (5) (or equivalently Figures 3 and 4) summarizes the effects of ‘bracket creep’ on tax rates. Except for the first jump in marginal rate (from 0 to 14%) the jumps in marginal rates are small at low income levels but become progressively larger as income increases. As displayed on Figures 3 and 4, treatment and control bands are roughly of the same size.

In column (6), I report the mean log difference of effective net-of-tax rates, $\log(1 - T'_2/1 - T'_1)$ for each group. Because individual real incomes change from year to year, figures in column (5) and (6) differ. The corresponding values are plotted on figure 5 for married taxpayers and Figures 6 and 7 for singles. The curve corresponding to column (5) is plotted in straight line while the curve corresponding to column (6) is plotted in dashed line. The curve of real changes in marginal rates goes up and down exactly in the same way as the curve of predicted changes in marginal rates.¹² Therefore, predicted change in marginal rates is highly correlated with the real change in rates and therefore predicted change is a good instrument for real change. However, because the spikes of real changes are flatter than the spikes of predicted changes, reduced form estimates similar to the ones previous studies report (see Section 2) would be significantly lower than structural estimates. I come back this point again later on.

In columns (7), (8) and (9), I report the means of log changes of real taxable income ($\log(\text{taxinc}_{2R}/\text{taxinc}_1)$), real adjusted gross income ($\log(\text{AGI}_{2R}/\text{AGI}_1)$) and real wages ($\log(\text{wages}_{2R}/\text{wages}_1)$). There is mean reversion at both ends of the income distribution. The change in incomes are high and positive for low incomes-this change is quickly decreasing as income increases-whereas the change in incomes becomes in general highly negative for high income earners. This complicates the estimation of the elasticities at very low and very high incomes.

If marginal rates matter for taxpayers, we should find that treatment groups experience larger decreases in incomes than the surrounding control groups. To check whether this pattern is apparent in the data, I have also plotted the log changes of taxable income and AGI on Figure 5 for married taxpayers and Figures 6 and 7 for ..

¹²The only exception is for the kink 18/19 for singles (see Figures 6 or 7) which is by far the smallest jump in the tax schedule. The predicted change does not follow as closely the real change for higher kink points (these kinks are not represented on the figures) because of the noise due to the small number of observations for high income earners.

singles.¹³

Figure 5 gives striking evidence of responsiveness of married taxpayers to tax rates. From the Treatment5 group (kink 21/24) to Control10 group (kink 43/49), the log change in taxable income presents exactly the same shape as the predicted changes in marginal rates: the value for the treatment group is always smaller than for the two surrounding control values. The same is true for log changes in adjusted gross income though the differences between treatments and controls are somewhat smaller. This is not the case for lower incomes because jumps in marginal rates were very small (less than 3%) except at the first kink (large jump of 14%). However, around this first kink, the mean reversion phenomenon is very important (this is not represented on Figure 5 but can be easily seen on Table I). Higher kink points do not reveal the same evidence but this may well be due to the small number of observations in that range¹⁴ and to mean reversion. The pattern of wage earnings¹⁵ is not similar to the pattern of taxable income or adjusted gross income: even at the middle income kinks, there is no clear evidence that wages of treatments tend to be systematically smaller than wages of surrounding controls. This already suggests that the response of taxpayers is probably not the consequence of reduced labor supply.

The pattern for singles on Figure 6 is less clear, even for middle income earners. Until Treatment8 group, the kinks were small (except the first one, the jumps were of less than 3%) and thus no systematic response is observed. From Treatment8 to Control12, there is some evidence of taxpayer behavior for adjusted gross income and taxable income. Above Control12, the number of observations becomes very small and no clear pattern would be observed. As for married taxpayers, wages for singles reveal no clear evidence of behavioral responses.

However, the first kink point for singles deserves particular attention. Figure 7 focuses more particularly on low income singles. There is a clear break in the pattern of AGI and wages around the first kink point consistent with a behavioral response to marginal rates: although the general pattern of the curves is declining (due to mean

¹³The log change in wages is also plotted on Figure 7.

¹⁴I have not plotted the curves for the highest kink points but this can be figured out looking at Table I.

¹⁵The curve for wages is not plotted to avoid packing too many plots on the figure.

reversion), wages and AGI go up from Treatment1 to Control1.¹⁶ There is no such pattern for taxable income because mean reversion in taxable income at the bottom is even larger than for AGI or wages.¹⁷ Therefore, Figure 7 suggests that low income singles reacted to marginal rates by reducing labor supply.

These figures suggest that taxpayers are responsive to changes in tax rates and that married taxpayers are more responsive to tax rates than singles. However, except for low income singles, wages do not seem to be responsive to changes in tax rates. I will now try to put numbers on these first qualitative results.

3.3.4 Wald Estimates

From the Tables described above, it is easy to compute Wald estimates of the elasticity for each kink. Wald estimates relate the difference in changes in income between treatments and controls to the difference in changes in real marginal rates between treatments and controls. This gives simple estimates of the elasticity of income with respect to marginal rates. Treatments are observations in a given treatment group and controls are observations belonging to the two surrounding control groups. The Wald estimate can be written as:

$$\hat{\zeta} = \frac{\hat{E}[\log(z_2/z_1)|Tr] - \hat{E}[\log(z_2/z_1)|C]}{\hat{E}[\log(1 - T'_2/1 - T'_1)|Tr] - \hat{E}[\log(1 - T'_2/1 - T'_1)|C]}$$

where \hat{E} means empirical mean, Tr is for treatment and C for control. z_1 is income in year 1 and z_2 is income in year 2 in terms of year 1 dollars. This estimate is equivalent to an IV regression of $\log(z_2/z_1)$ on $\log[(1 - T'_2)/(1 - T'_1)]$ (and a constant) using a binary instrument (1 if in treatment and 0 if in control). This method leads to consistent estimates if the difference in changes in income between treatments and controls is entirely due to the fact that treatments are more likely to experience an increase in rates than controls. This assumption is likely to be satisfied because incomes of treatments and surrounding controls are very close and therefore treatments

¹⁶Wages and AGI curves are very close for low income singles because most of them report only wage income.

¹⁷This is explained by the deduction of exemptions and net itemized deductions from AGI to compute taxable income: this overstates mechanically positive log changes in taxable income compared to AGI for very low income earners.

and controls are similar except for their treatment/control status.¹⁸ The IRS panel does not contain covariates (such as age or educational attainment) which could have been used to test formally whether Treatments and surrounding Controls are similar. Though not attempted in this paper, it would be possible to use another dataset with many covariates (such as the CPS or the PSID) and define the income groups corresponding (roughly) to Treatment and Control groups so as to test formally whether they are the same.

Reduced form estimates can also be derived by simply running an OLS regression of $\log(z_2/z_1)$ on $\log(1 - T'_p/1 - T'_1)$ (and a constant). This corresponds exactly to the methodology used by Navratil (1995).

I have reported Wald and Reduced form estimates for middle income kinks for taxable income, Adjusted Gross Income (AGI) and wages for married taxpayers and singles on Table III. Each Wald estimate was computed using observations of the corresponding treatment and both surrounding controls.

Columns (1) and (2) display the location of kink points (in current dollars) and the corresponding jumps in marginal rates. Column (3) presents the difference in the change in taxable income between the treatment group and the two surrounding control groups. This difference can be derived directly from Tables I and II using column (6) (which gives the average change in taxable income for each treatment and control group) and column (7) (which gives the sample weights for each treatment and control group).¹⁹ Column (4) gives the values of the instrument $\log[(1 - T'_p)/(1 - T'_1)]$ for each Treatment group.

Column (5) presents the difference of the log change in marginal rates ($\log[1 - T'_2/1 - T'_1]$) between the treatment group and the two surrounding control groups.²⁰ Column (6) presents the reduced form estimates: this is just column (3) divided by column (4). Column (7) presents the Wald estimates (this is column (3) divided by column (5)). Wald estimates for adjusted gross income and wages have been computed in the same way (the different steps are not reported) and are presented in columns

¹⁸This assumption is much more likely to be satisfied for middle income earners where mean reversion is not an issue. That is why I give Wald estimates only for middle income kinks.

¹⁹For example, the first number 0.0006 in Column (3) of Table III is obtained as $[(-0.0691 \cdot 2241) - (-0.0665 \cdot 2991 - 0.0727 \cdot 3264)]/[2241 + 2991 + 3264]$.

²⁰This difference can be derived from Tables I and II exactly in the same way as Column (3) of Table III.

(8) and (9). All standard errors have been computed by running the corresponding OLS (for the reduced form estimates) and IV (for the Wald estimates) regressions.

I have tried alternative estimates. Removing taxpayers at the frontier between control and treatment bands did not change much the estimates. Keeping only the controls close to the treatments (i.e. discarding the controls which are the further away from the treatments) did not modified much my results either.

Looking at Table III, we can note that the elasticities are in general positive; this means that treatments tend to experience larger decreases in income than controls and thus that taxpayers are responsive to marginal rates. The estimates confirm the patterns of Figures 5 and 6: the estimates are significant and large for married middle income earners for taxable income and AGI. The estimates are in general larger for taxable income than for AGI. The estimates for wages are usually much more smaller, often very near 0. As pointed out before, the estimates for singles are lower and not significant. The reduced form estimates are equal to about one-half of the structural estimates.

Therefore, simple Wald estimates confirm our first qualitative results. The response is higher for married taxpayers than for singles. The response of taxable income is higher than AGI and especially than wages. The response of wages is almost never significantly different from 0. However, the results are not estimated with great precision and there is large variability across kink points. The aim of next sections is to compute estimates based on larger portions of the income distribution in order to obtain more precise results.

3.4 Model and Identification Strategy

This section uses a regression framework to aggregate estimates over several kink points. A simple model will illustrate the issues at hand and show that the estimated elasticities are in fact pure compensated elasticities. The budget constraint of a taxpayer on a linear part of the tax schedule is given by $c = z(1 - \tau) + R$, where z represents before tax income, τ is the marginal rate and R is virtual income. The virtual income R is the post-tax income that the taxpayer would get if he reported no income and was allowed to stay on the same budget set line (with constant marginal

rate τ). From individual utility maximization, we can derive a income supply function which depends on the slope of the budget line and on virtual income.

$$z = z(1 - \tau, R)$$

From this income supply function, the uncompensated elasticity of income (denoted by ζ^u) and income effects (denoted by η) can be defined as follows:

$$\zeta^u = \frac{1 - \tau}{z} \frac{\partial z}{\partial(1 - \tau)}$$

and,

$$\eta = \frac{\partial z}{\partial R}$$

Let $z^c = z^c(1 - \tau, u)$ be the compensated income supply.²¹ The compensated elasticity of income (ζ^c) is defined by:

$$\zeta^c = \frac{1 - \tau}{z^c} \frac{\partial z^c}{\partial(1 - \tau)}$$

The two elasticities and income effects are related by the Slutsky equation:

$$\zeta^c = \zeta^u - (1 - \tau)\eta \quad (3.1)$$

'Bracket creep' can be seen as a change in both virtual income R and marginal rate τ . Small changes in R and τ affect income supply z as follows,

$$dz = -\frac{\partial z}{\partial(1 - \tau)}d\tau + \frac{\partial z}{\partial R}dR$$

Using the definition of elasticities, we get:

$$dz = -\zeta^u z \frac{d\tau}{1 - \tau} + \eta dR$$

Using the Slutsky equation (3.1) and rearranging,

²¹ $z^c(1 - \tau, u)$ is the income supply which minimizes costs to attain utility level u for a given tax rate τ .

$$\frac{dz}{z} = -\zeta^c \frac{d\tau}{1-\tau} + \eta \frac{dR - zd\tau}{z}$$

To introduce randomness in the model, I suppose that the income supply function z also shifts randomly (i.e. $dz/z = \epsilon$) from year to year for reasons unrelated with the tax change. The random variable ϵ can be considered as taste shocks (resulting for example from a change in the composition of the household) or random changes in work opportunities (such as unexpected unemployment or job change). Therefore, the equation giving the total change in income from year 1 to year 2 (dz/z) can finally be written as:

$$\frac{dz}{z} = -\zeta^c \frac{d\tau}{1-\tau} + \eta \frac{dR - zd\tau}{z} + \epsilon \quad (3.2)$$

Let us first neglect the income effect term (i.e., assume that $\eta = 0$). In that case, by the Slutsky equation (3.1), compensated and uncompensated elasticities are the same (I note $\zeta = \zeta^c = \zeta^u$). Assuming that changes from year to year are small, we have, $dz/z \simeq \log(z_2/z_1)$ and $-d\tau/(1-\tau) \simeq \log[(1-T'_2)/(1-T'_1)]$ (with the same notation as in the previous section). The corresponding regression framework would then be the following:

$$\log(z_2/z_1) = \zeta \log[(1-T'_2)/(1-T'_1)] + \epsilon$$

Now clearly, $\log[(1-T'_2)/(1-T'_1)]$ is correlated with the error term because if ϵ is large, income goes up and thus, because marginal tax rates are increasing with income, $1-T'_2$ decreases. Therefore an OLS regression leads to estimates badly biased downward.²² However, it is possible, using the variation in tax rates due to ‘bracket creep’, to construct instrumental variables. Consider the following dummy variable,

$$instr_{is} = 1(\text{taxinc}_1 \in \text{Treatment for Kink } i, \text{mars} = s)$$

These are binary instruments equal to 1 exactly for taxpayers whose taxable income in year 1 (denoted by taxinc_1) is in the treatment for Kink i and whose marital

²²In fact, running OLS regressions always leads to elasticity estimates well below zero (in general below -3).

status is s . The marital status mar_s can take two values: 0 for singles and 1 for married taxpayers. The instruments $instr_{i,s}$ depend only on the level of income in year 1. Therefore, in this simple model, the instruments depend only on z_1 and are uncorrelated with ϵ if ϵ is independent of z_1 . In this case we would just have to run the following regression:

$$\log(z_2/z_1) = \zeta \log[(1 - T'_2)/(1 - T'_1)] + \epsilon \quad (3.3)$$

using $instr_{i,s}$ as instruments for the real variation in marginal rates. The elasticity parameter ζ would be estimated consistently. Note that this set-up leads exactly to the simple Wald estimates presented above where we restricted ourselves to small portions of the distribution of income so that only one instrument was used for each regression.

However, if we consider large portions of the income distribution, it is more realistic to assume that the size of the random change in incomes (i.e ϵ) varies as we move along the distribution of income. We have seen in the previous section that there is mean reversion and therefore that if z_1 is low in year 1 then z_2 is very likely to be above z_1 . In this case, the distribution of the random shock in income ϵ is likely to be skewed toward the right. This works in the other direction for high income earners in year 1. On the other hand, if there is an underlying increase in inequalities (i.e., the rich get richer and the poor get poorer), a component of ϵ will be positively related to income in year 1 because high income earners will tend to do even better whereas low income earners will tend to do worse.

So if ϵ depends on $z_1 = taxinc_1$, the instrument (which is also a function of z_1) is likely to be correlated with the error term ϵ . However by controlling for any smooth function of $taxinc_1$ in the regression set-up in both stages, it is possible to get rid of the correlation between ϵ and the instruments. The parameter of interest remains identified as long as the dependence of ϵ with respect to $taxinc_1$ does not reproduce the shape of the instruments. This dependence is due to mean reversion, macro-economic shocks and underlying trends in the income distribution and therefore is probably very smooth compared to the dummy shape of the instruments. Therefore,

the system is very likely to be well identified.²³ Note that previous tax reform studies (which were reviewed in Section 2) cannot control for income because the marginal cuts were increasing in income (thus their instrument is monotone in income) and therefore controlling for income would destroy the identification.

Let us now analyze the case with income effects in equation (3.2). $dR - z d\tau$ is the change in after-tax income due to the tax change for a given before tax income z : this is the vertical distance between the tax schedule for year 1 and the tax schedule for year 2 on Figure 1. This quantity is *continuously* increasing in income²⁴ and thus affects treatments and controls in roughly the same way. Therefore, this additional income effect term can be incorporated in the error term. The dependence of this term on income will be controlled for by the functions in $taxinc_1$ included as controls in the regression. Therefore, even with income effects, the parameter ζ that I estimate is in fact the compensated elasticity ζ^c . Intuitively, at a given kink point, the increase in tax liability due to ‘bracket creep’ is roughly the same for treatments and controls but the change in tax rates is different for the two groups. Therefore, the difference in behavioral responses between the two groups is due to pure substitution effects. Thus, the ‘bracket creep’ experience allows the estimation of a conceptually well defined parameter. This point is important because the tax reform studies reviewed in Section 2 were only able to identify elasticity estimates which were a mix of substitution and income effects.

Let me now describe precisely the regression framework and the covariates I will use. To allow more generality, I run regressions in levels: $\log(z_2)$ is the dependent variable instead of $\log(z_2/z_1)$ and I include $\log(z_1)$ in the list of controls on the right hand side. When I run a regression for both married taxpayers and singles, I add a dummy $mars$ for marital status ($mars$ is equal to one if married and zero if single). I also add a dummy $item$ for being an itemizer in year 1. Being an itemizer in year 1

²³This strategy is conceptually close the Regression Discontinuity Design (RDD), used in Angrist and Lavy (1999) and Van der Klauw (1996). The idea in both papers is to use the fact that the treatment (class size for Angrist and Lavy, financial help decision for Van der Klauw) is assigned on the basis of a discontinuous function of a continuous variable. The strategy is to use the rule as a source of identification, controlling in the regression for smooth functions of the variables on which the selection is based.

²⁴This quantity is not increasing smoothly because, as displayed on Figure 1, it is constant over Control regions and linearly increasing over Treatment regions. However, the important point here is that the quantity does not jump discontinuously.

is predetermined and therefore *item* can be considered as an independent variable.²⁵ Therefore the specification is as follows:

$$\log(z_2) = \alpha_0 + \alpha_1 \log(z_1) + \alpha_2 mars + \alpha_3 item + \zeta^c \log[(1 - T'_2)/(1 - T'_1)] + \beta f(taxinc_1) + \epsilon \quad (3.4)$$

The first stage being:

$$\log[(1 - T'_2)/(1 - T'_1)] = \sum_{i,s} \gamma_{is} instr_{is} + \pi_0 + \pi_1 \log(z_1) + \pi_2 mars + \pi_3 item + \delta f(taxinc_1) + \nu \quad (3.5)$$

where:

- z_i is real income in year i (this can be taxable income, wages or AGI),
- T'_i is the marginal rate in year i (i.e. T' (nominal taxable income)),
- ζ^c is the parameter of interest: compensated elasticity of income with respect to marginal rates.

The controls $f(taxinc_1)$ are smooth functions of $taxinc_1$ (polynomials in $taxinc_1$). Polynomials are added until the elasticity estimate is stabilized (3 or 4 polynomials are enough in most cases).

An alternative would be to define a single instrument: $\log(1 - T'_p/1 - T'_1)$ for all the regressions (Auten and Carroll (1997) used this type of instrument in their study). This single instrument would impose a relation between the size of the jump in marginal rates and the value of the instrument. Results with a single instrument are very similar to the results I present. Increasing the number of instruments, however, increases the power of the first stage and therefore reduces a little bit the standard errors, that is why I choose the multi-instrument set-up. I do not use exactly one instrument for each kink because some low kinks are very small and I have few observations for the highest kinks. Therefore I have grouped some kinks together to avoid using too weak instruments. The precise grouping is described in appendix.

²⁵I add the dummy *item* because I show below that the elasticities of taxpayers itemizing in year 1 are significantly higher than the elasticities of non-itemizers.

3.5 Regression results

The first stage always leads to very significant coefficients for all the binary instruments. The F-statistic for the joint test of all the coefficients of the instruments being null is always higher than 50. This confirms that the instruments are good in the sense that they are significantly correlated with the endogenous regressor.

I estimated equation (3.5) for three types of incomes: wages, adjusted gross income (AGI) and taxable income and different portions of the income distribution.²⁶ I divided my sample according to marital status - Single taxpayers and Married taxpayers filing jointly - and into year 1 itemizers and year 1 non-itemizers and estimated elasticities for those different groups. I did not split the sample of singles into itemizers versus non itemizers because very few singles choose to itemize and thus estimates would have been fairly imprecise for that sub-group. The elasticity results are presented in Tables IV and V.²⁷ Table IV presents estimates for a wide range of incomes (columns (1) to (3)) and for middle income earners (columns (4) to (6)). Table V focuses on high income earners (columns (1) to (3)) and on low income earners around the first kink point (columns (4) and (5)).

Column (1) of Table IV suggests that elasticities of taxable income are smaller than those found in previous studies using tax reforms: around 0.3 for married taxpayers and singles together, around 0.4 for married taxpayers and around 0.2 for singles. The elasticities of adjusted gross income are slightly lower: around 0.2 (see column (2)). The elasticities of wages are even smaller (around 0.1). The elasticities are in general higher for married taxpayers than for singles. Note however that the elasticities are not estimated with very high precision and therefore most of the estimates are not significantly different from 0. The estimated elasticities suggest that the labor supply response to marginal rates is small. This is consistent with the estimates of traditional labor supply literature.

The most striking fact in Table IV is that the elasticity for non-itemizers is always much smaller (and often slightly negative) than the elasticity of itemizers. Elasticity

²⁶All income levels are expressed in 1979 dollars; a dollar of 1979 corresponds to 2.3 dollars of 1997.

²⁷In both Tables, the list of polynomial controls in *taxinc*₁ is reported in the note. $\log(z_1)$ is always included as a covariate in the regressions.

for married itemizers are high and significant: 0.65 for taxable income and 0.4 for adjusted gross income. The difference between the elasticity estimates of itemizers and those of non-itemizers persists for adjusted gross income and wages, though it is in general smaller than the difference for taxable income. This means that itemizers react more than non-itemizers not only through an increase in their itemized deductions but also through a larger reduction in reported income. This suggests that the population of itemizers is different from the population of non-itemizers. The possibility of itemizing plays the role of a screening device where elastic taxpayers choose to itemize and non-elastic taxpayers choose the standard deduction.

Columns (4), (5) and (6) of Table IV report the same kind of estimates but restricted to middle income earners. The general pattern is the same as in columns (1) to (3). However, the elasticities for this group are, in general, significantly higher than for the wider range of income: 0.4 for taxable income, 0.3 for adjusted gross income for married taxpayers and singles together, 0.7 for taxable income of married itemizers. Note that this high value is close to the results of Navratil (1995) and Auten and Carroll (1997). The wage elasticity of married taxpayers, which is around 0.3, is also somewhat higher than before.

Table V focuses more specifically on high income earners (columns (1), (2) and (3)) and on low income earners around the first kink point (columns (4) and (5)). The elasticities of high income earners are smaller than those of middle income earners: around 0.3 for taxable income, around 0 for adjusted gross income and negative (though never significant) for wages. The elasticities, however, are not estimated with very high precision. This seems to indicate that high income earners did not react as much as middle income earners to 'bracket creep'. The discrepancy between the results for adjusted gross income and taxable income probably means that most of the response of high income earners was through increased itemized deductions and not through a reduction in real earnings.

Columns (4) and (5) in Table V report estimates around the first kink point. The estimates confirm our previous qualitative results in Section 3. The elasticity of adjusted gross income and wages is large and significant for singles: 1.1 for adjusted gross income and 1.3 for wages. These are the largest elasticities found in this study. This suggests that the elasticity of labor supply is potentially high for singles with

low incomes. Low income earners have few possibilities of altering their tax liabilities through a change in reporting behavior and therefore the decrease in reported wages is likely to be the consequence of reduced labor supply. Note however that elasticities of low income earners can be high even if the response to taxation is small in absolute levels. This is due to the fact that the elasticity measures the response relative to the size of income (which is small for low income earners).

The elasticity is about 0 (even slightly negative) for married taxpayers. Wald estimates would not have shown accurate results because mean reversion is important in the low end of the income distribution and therefore it is important to control for income.²⁸ The mode of the income distribution is slightly on the left of the first kink point for singles and many singles have their permanent income around this point and are likely to react to taxes at this level. Note also that tax liabilities begin at the first kink point and therefore taxpayers may perceive more accurately this jump in marginal rates than those of other kink points. This may partly explain the high response of singles around this point. Mean reversion is stronger for married taxpayers because many low income married taxpayers are only transitorily around that point and are less likely to react to ‘bracket creep’ than singles.

The estimates shown in Tables IV and V broadly confirm the results of Section 3 where we noticed that married middle income earners are the most responsive but that the response of low income singles was also significant. Except for this last group, the response of wages is small, therefore income response to marginal rates may be due to changes in reporting behavior rather than reduced labor supply. Most of the response comes from the population of itemizers who is more elastic and can partly decrease its tax liability through increased itemized deductions.

3.6 Conclusion

This paper has made an attempt at identifying the impact of marginal rates on various types of reported income using ‘bracket creep’ as a source of variation in tax rates.

²⁸Plugging too many covariates is not possible either, because there is only one instrument in these regressions and too many covariates would destroy the identification. I have therefore included only two covariates: $\log(z_1)$ and $taxinc_1$.

The particular nature of this tax change allowed me to divide the sample between treatments and controls over the whole range of income distribution. Therefore, the estimates presented are not biased by possible underlying trends in income distribution such as mean reversion or a rise in income inequality. Most results point to the general conclusion that there is a response of taxpayers to tax rates: incomes of taxpayers in the treatment groups tend to decrease more than incomes of taxpayers in the control groups. Moreover, the estimates are somewhat higher than traditional labor supply estimates but smaller than those found in previous studies using tax reforms. The estimates are in general higher for married taxpayers than for singles and higher for itemizers than for non-itemizers. Moreover, the estimates are in general higher for taxable income than for adjusted gross income and higher for adjusted gross income than for wages. This suggests that most of the response is due to changes in reporting behavior rather than reduced labor supply. Except for singles at the bottom of the income distribution,²⁹ wage elasticity estimates found in this study are very small and comparable to the estimates found in most labor supply studies. Part of the higher elasticities of married taxpayers compared to singles may be due to the higher responsiveness of secondary earners to tax rates, which is well documented in the literature. Indeed, wage elasticity estimates for married taxpayers are almost always higher than the estimates for singles.

Three caveats should be mentioned. First, my study captures only short term effects of marginal rates because it compares outcomes only across consecutive years. If responses to marginal rates are slow, my estimates may be smaller than medium or long term elasticities. However, several studies about behavioral responses to taxation suggest that short term responses are likely to be higher than long-term responses. Slemrod (1995) argues that the timing of economic transactions is the most responsive to tax incentives (the response of real economic activities seems to be much lower). Goolsbee (1997), using a panel data on corporate executive compensation, showed that the income tax increase of 1993 led to large short term inter-temporal income shifting but that the long term response was small. In the 'bracket creep' experience, as inflation was expected, there may also be an inter-temporal substitution effect.

²⁹In any case, this result must be considered with caution because it is based on behavior around a single kink point.

People know that taxes will be higher in the following year and therefore try to increase their income now at the expense of next year's income. Moreover, after Reagan's election in 1980 people knew that taxes would be cut by 1982. This gave another incentive to shift income away from years 1980 and 1981. However, this expected reduction in taxes probably affected treatments and controls in the same way and therefore my estimates are not affected by this expectation component.

Second, as 'bracket creep' was not a tax reform, taxpayers may not have been fully aware of the marginal tax increases and thus did not respond to the change. This seems unlikely because 'bracket creep' was perceived as a major income tax event which triggered what has been called the 'tax revolt' of the late 1970s and early 1980s. If we assume that only a part of all taxpayers were aware of the effects of 'bracket creep', then the responses I measure are due only to these 'aware' taxpayers. The elasticity estimates for these taxpayers would therefore be equal to my estimates divided by the proportion of 'aware' taxpayers. However, to get elasticity estimates as high as those found in previous tax reform studies, the proportion of 'aware' taxpayers should have been unrealistically low.

Last, my study measured the response of relatively small changes in tax rates and found smaller elasticity estimates than previous studies. It may be the case that the response for larger tax rates cuts (such as ERTA or TRA) cannot be directly predicted from the results presented here. In other words, responses of taxpayers may be non-linear: a small change can lead to almost no effect while a big change can have a dramatic impact on reported income.³⁰

Despite these caveats, the present study using 'bracket creep' has important advantages over studies exploiting tax reforms and has taught us interesting facts about the behavioral responses to marginal tax rates. In future work, I plan to develop the model presented in Section 4 in order to derive a general method to estimate both income and substitution effects using panel data on tax returns and several tax reforms at the same time. The method would be less dependent on structural form assumptions than most labor supply studies because the identification would come di-

³⁰This is probably what happened after the TRA of 1986 for very rich taxpayers who have the possibility to change the way and the timing in which they report income. See Feenberg and Poterba (1993) and Slemrod (1996). This non-linear behavior is probably much less relevant for low and middle income earners.

rectly from tax reforms. The methodology would however impose more structure than previous tax reform studies to allow the estimation of both income and substitution effects.

Appendix

Computations of Deflated Taxable Income and Predicted Taxable income:

I denote by *exempt* the level of exemptions, by *stded* the level of the standard deduction and by *itemz* the nominal level of itemized deductions in year 2. From taxable income in year 1 ($taxinc_1$), I compute predicted taxable income ($taxinc_p$) which is $taxinc_1$ expressed in year 2 dollars. I assume that nominal adjusted gross income and nominal itemized deductions grow at the inflation rate denoted by π . Nominal exemptions and standard deductions stay constant, therefore for non-itemizers, we have,³¹

$$taxinc_p = AGI_p - exempt = (1 + \pi)AGI_1 - exempt = (1 + \pi)taxinc_1 + \pi exempt$$

For itemizers, we have,

$$taxinc_p = (1 + \pi)AGI_1 - exempt - (1 + \pi)item + stded = (1 + \pi)taxinc_1 + \pi(exempt - stded)$$

We now have to express the value of $taxinc_2$ in year 1 dollars. Again, we have to take into account the fact that *exempt* and *stded* are not indexed, therefore we compute real taxable income in year 2 (denoted by $taxinc_{2R}$) as follows:

$$taxinc_{2R} = \frac{AGI_2}{1 + \pi} - exempt = \frac{taxinc_2}{1 + \pi} - \frac{\pi exempt}{1 + \pi}$$

for non-itemizers in year 2.

$$taxinc_{2R} = \frac{AGI_2}{1 + \pi} - exempt - \frac{itemz_2}{1 + \pi} + stded = \frac{taxinc_2}{1 + \pi} - \frac{\pi(exempt - stded)}{1 + \pi}$$

³¹Note that because of inflation, a non-itemizer may become an itemizer if his potential itemized deductions are just below the standard deduction. This would change $taxinc_p$ by a small amount ($\pi stded$) and we thus neglect this possibility.

for itemizers in year 2 such that³² $itemz_2/(1 + \pi) \leq stdded$.

$$taxinc_{2R} = \frac{taxinc_2}{1 + \pi} - \frac{\pi exempt}{1 + \pi} + \frac{itemz_2 - stdded}{1 + \pi}$$

for itemizers in year 2 such that³³ $itemz_2/(1 + \pi) < stdded$.

Description of the grouping of instruments:

To avoid using too weak instruments, I have grouped the instruments for each kink as follows:

- (14/16 and 16/18), (18/19 and 19/21), (21/24 and 24/26), (34/39 and 39/44), (44/49 and 49/55), (55/63 and 63/68 and 68/70) for singles
- (14/16 and 16/18), (18/21 and 21/24), (43/49 and 49/54), (54/59 and 59/64 and 64/68 and 68/70) for married taxpayers.

Therefore I have 9 instruments for each marital status instead of 15 for married taxpayers and 16 for singles. When I have grouped several kinks, I have given values proportional to $\log(1 - T'_p/1 - T'_1)$ for each kink and thus some instruments are no longer binary but can take 3 to 5 different values. This grouping device does not noticeably affect the results and avoids using too weak instruments.

³²That is, taxpayers whose itemized deductions are large enough so that even with deflated itemized deductions, it is still advantageous to itemize in year 1.

³³That is, taxpayers whose itemized deductions are just above the standard deduction so that deflating the itemized deductions makes itemizing unattractive in year 1.

Figure 1: Bracket Creep Experiment, Tax Schedules

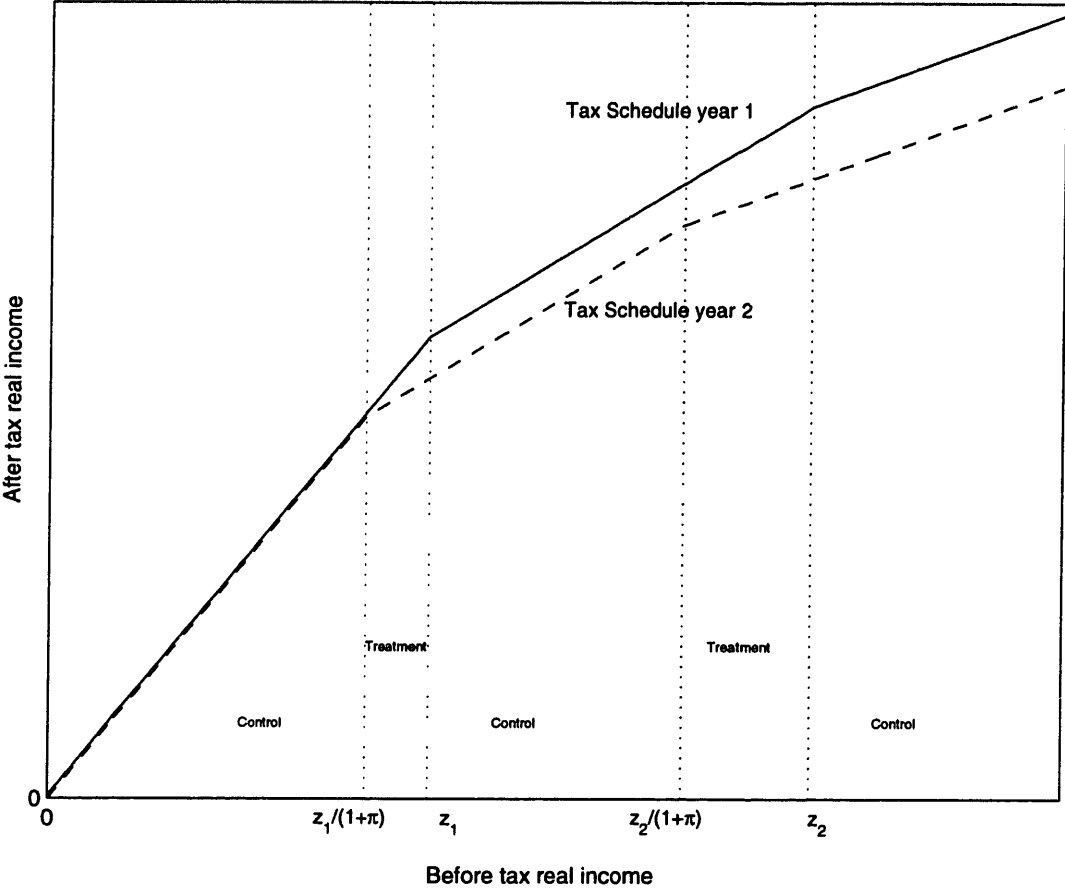


Figure 2: Bracket Creep Experiment, Marginal Rates

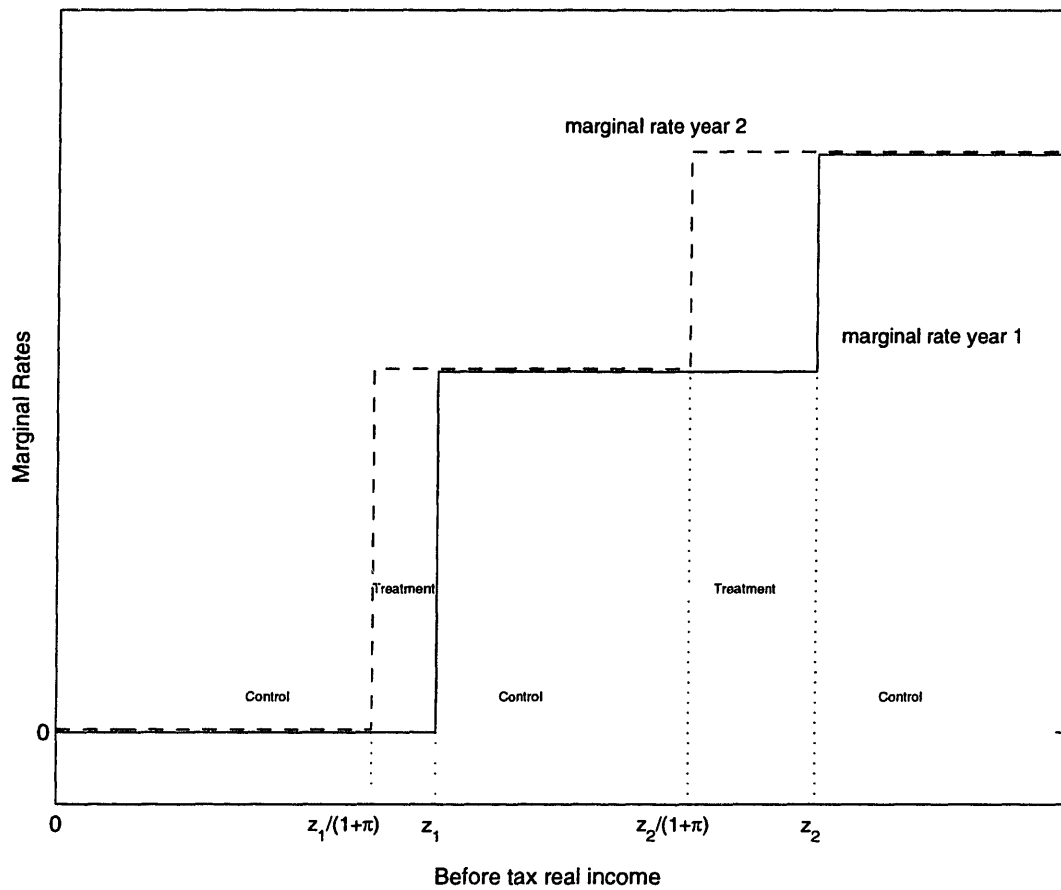


Figure 3: Shift in Marginal Rates for Married Taxpayers

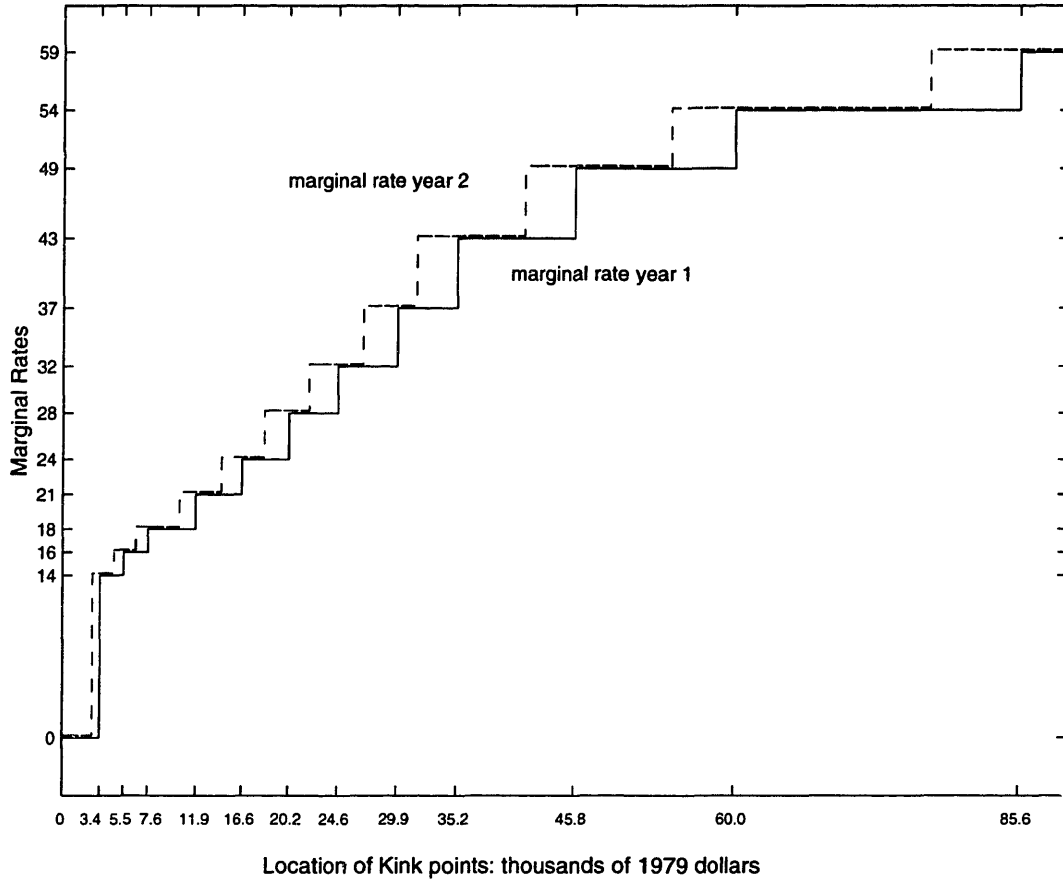


Figure 4: Shift in Marginal Rates for Single Taxpayers

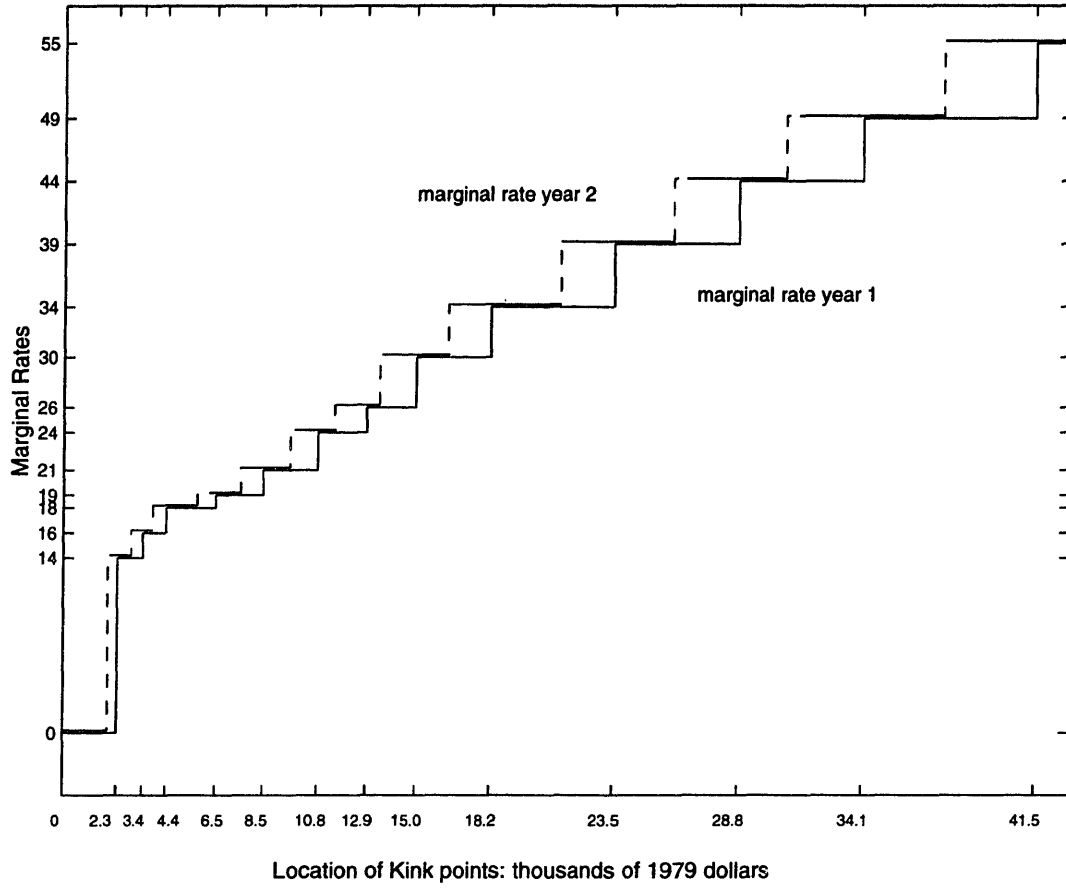


Figure 5: Married Taxpayers

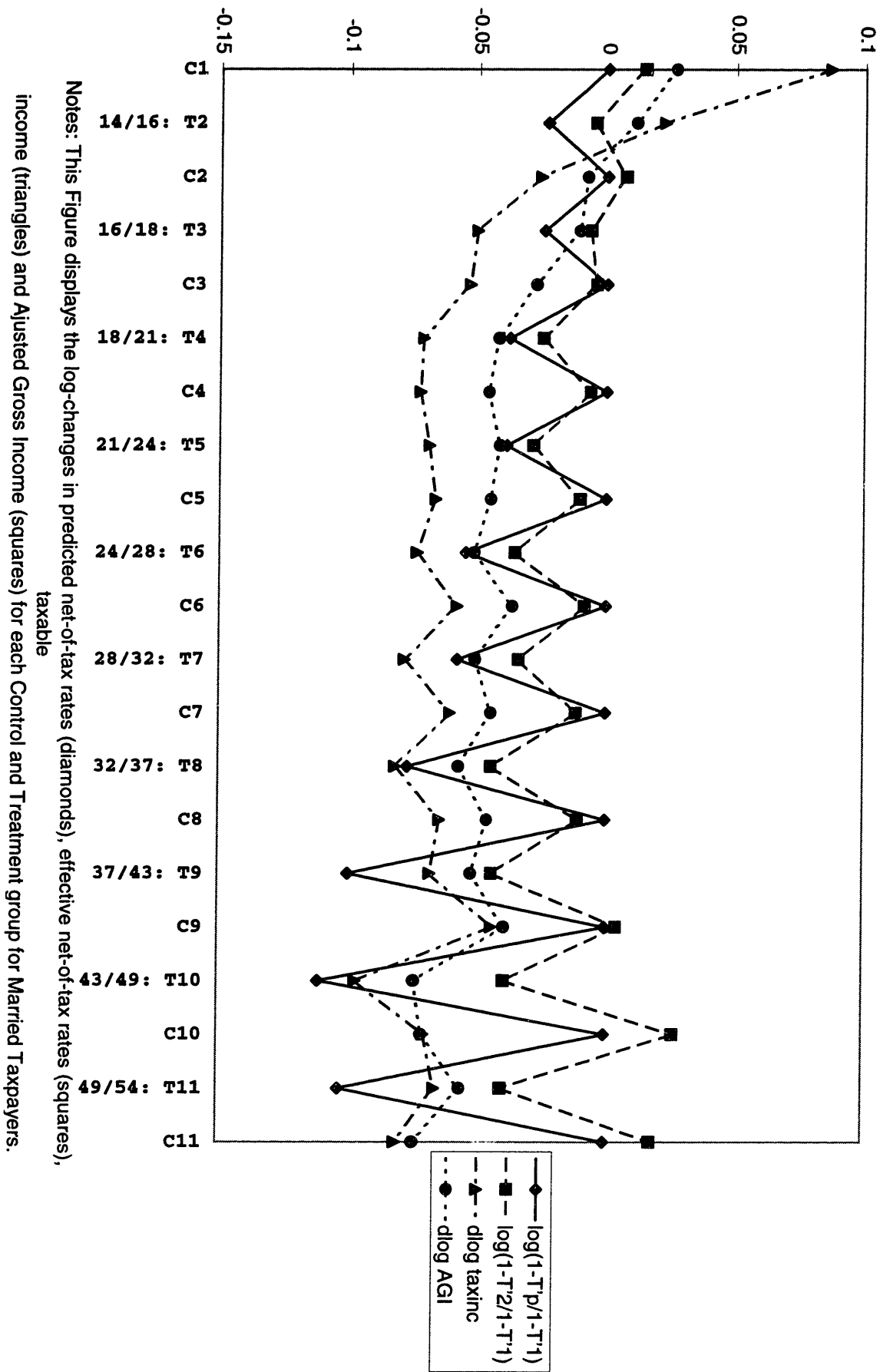
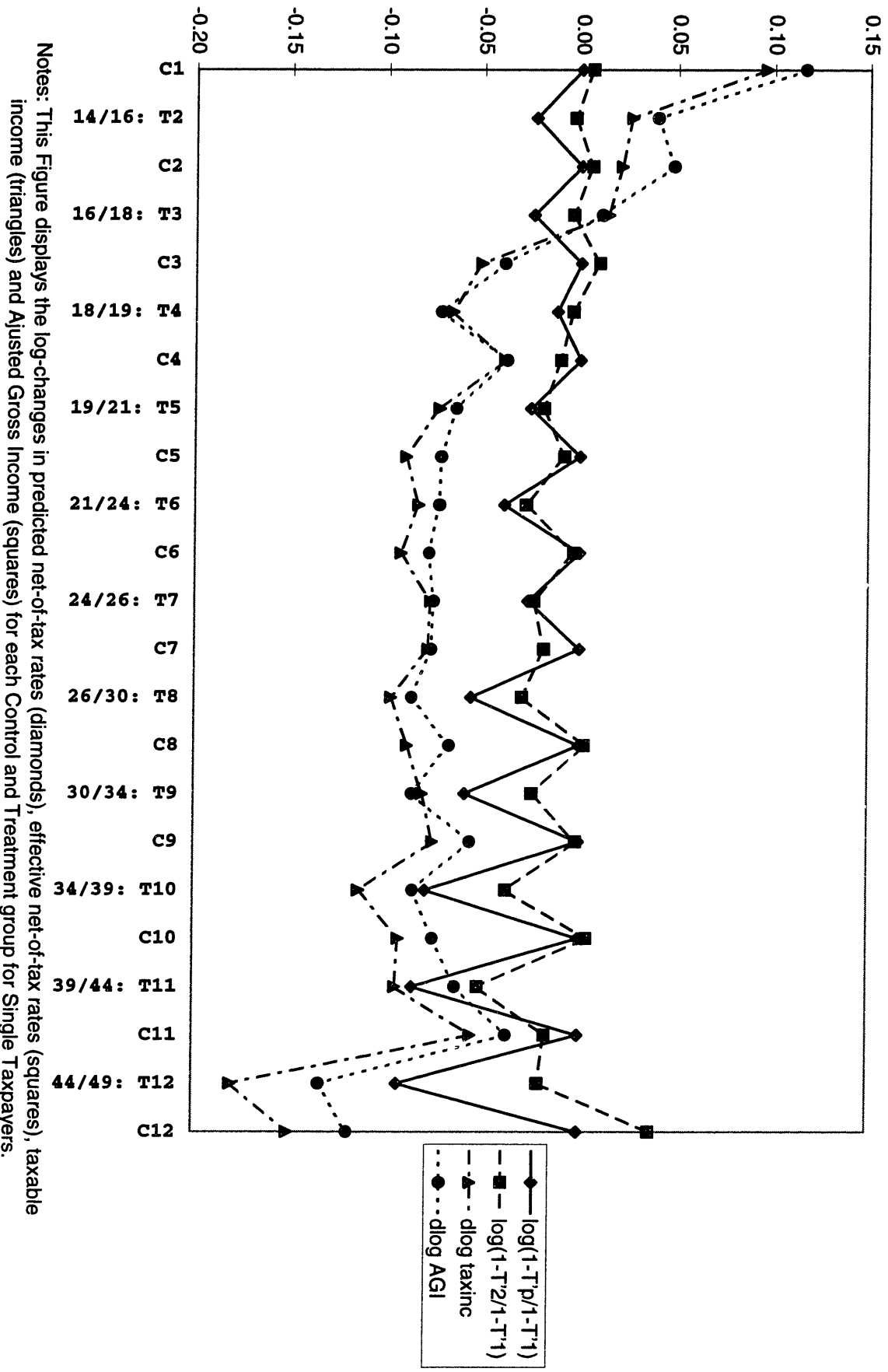
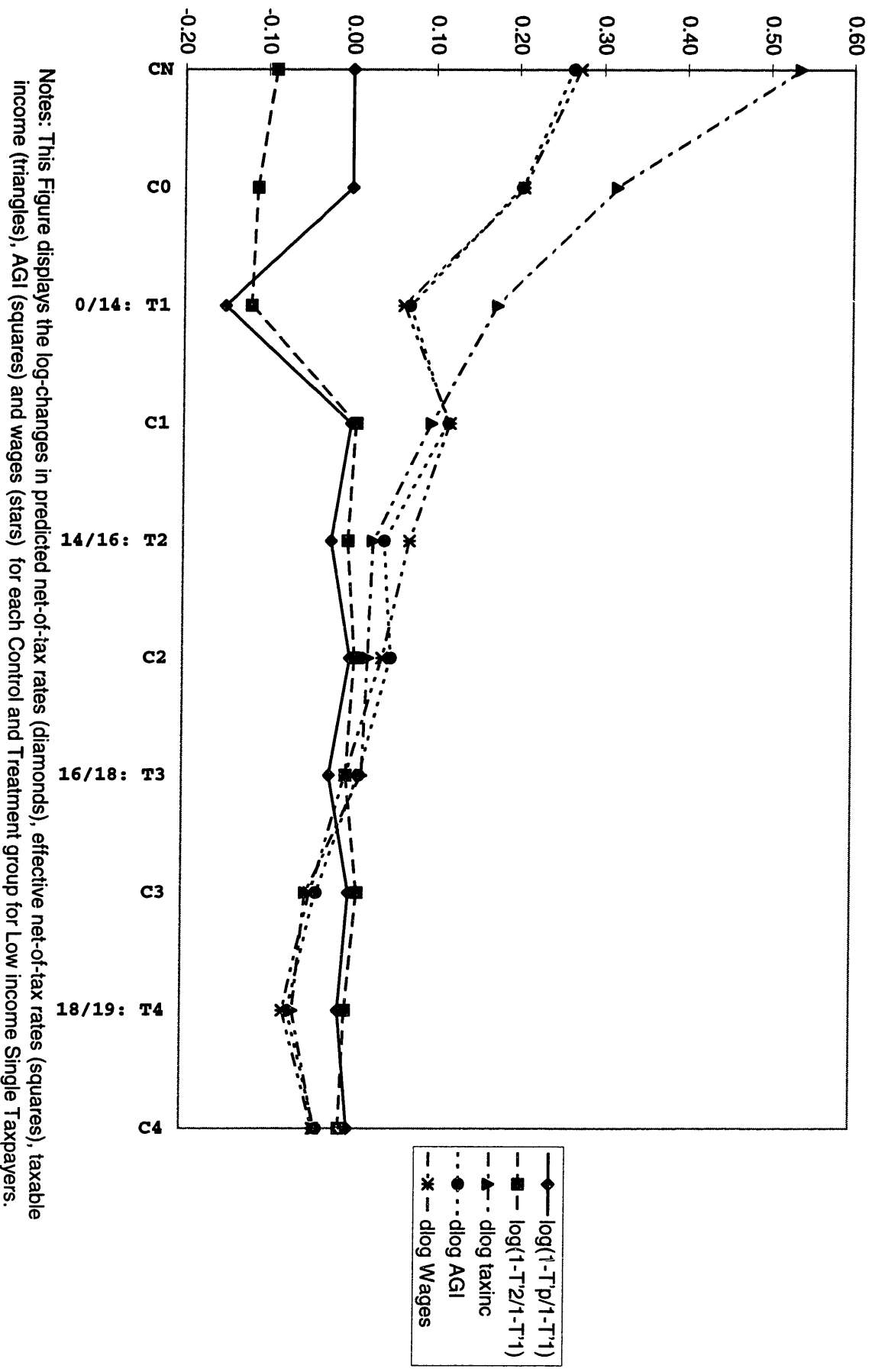


Figure 6: Single Taxpayers



Notes: This Figure displays the log-changes in predicted net-of-tax rates (diamonds), effective net-of-tax rates (squares), taxable income (triangles) and Adjusted Gross Income (squares) for each Control and Treatment group for Single Taxpayers.

Figure 7: Singles, low income earners



Notes: This Figure displays the log-changes in predicted net-of-tax rates (diamonds), effective net-of-tax rates (squares), taxable income (triangles), AGI (circles) and wages (stars) for each Control and Treatment group for Low income Single Taxpayers.

Table 1: Summary statistics for Married Taxpayers

Groups	Kinks			Number of observations	log(1-T ^p /1-T ¹)		dlog of taxable income		dlog of adjusted gross income		dlog of wages	
	Location	Jump	(1)		(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Control N				392	0	-0.0723	0.684	0.131	0.103			
Control 0				852	0	-0.0861	0.340	0.0859	0.069			
Treatment 1	\$3,400	0/14		605	-0.1508	-0.1134	0.141	0.0700	0.0537			
Control 1				1,383	0	0.0143	0.0865	0.0262	0.0180			
Treatment 2	\$5,500	14/16		783	-0.0233	-0.0049	0.022	0.0111	-0.0185			
Control 2				1,222	0	0.0072	-0.0258	-0.00797	-0.0265			
Treatment 3	\$7,600	16/18		1,084	-0.0243	-0.0066	-0.0506	-0.0109	-0.0307			
Control 3				3,550	0	-0.0043	-0.0533	-0.0275	-0.0536			
Treatment 4	\$11,900	18/21		1,615	-0.0377	-0.0247	-0.0716	-0.0420	-0.0542			
Control 4				3,264	0	-0.0065	-0.0727	-0.0458	-0.0531			
Treatment 5	\$16,000	21/24		2,241	-0.0387	-0.0284	-0.0691	-0.0414	-0.0545			
Control 5				2,991	0	-0.0104	-0.0665	-0.0448	-0.0553			
Treatment 6	\$20,200	24/28		2,580	-0.0545	-0.0354	-0.0737	-0.0513	-0.0613			
Control 6				2,294	0	-0.0093	-0.0578	-0.0364	-0.0560			
Treatment 7	\$24,600	28/32		2,230	-0.0576	-0.0338	-0.0782	-0.0508	-0.0470			
Control 7				1,908	0	-0.0116	-0.0605	-0.0445	-0.0535			
Treatment 8	\$29,900	32/37		1,634	-0.0769	-0.0444	-0.0816	-0.0571	-0.0797			
Control 8				883	0	-0.0109	-0.0644	-0.0459	-0.0487			
Treatment 9	\$35,200	37/43		971	-0.0998	-0.0440	-0.0681	-0.0521	-0.0669			
Control 9				1,057	0	0.0045	-0.0439	-0.0390	-0.0723			
Treatment 10	\$45,800	43/49		418	-0.1109	-0.0390	-0.0968	-0.0741	-0.0754			
Control 10				339	0	0.0264	-0.0703	-0.0712	-0.0797			
Treatment 11	\$60,000	49/54		151	-0.1031	-0.0400	-0.0659	-0.0560	-0.0945			
Control 11				195	0	0.0179	-0.0809	-0.0742	-0.0685			
Treatment 12	\$85,600	54/59		67	-0.1154	-0.0794	0.0097	0.0413	0.142			
Control 12				73	0	0.0554	-0.1137	-0.0812	0.0167			
Treatment 13	\$109,400	59/64		31	-0.1301	-0.0374	-0.0790	-0.0430	0.591			
Control 13				102	0	0.0388	-0.111	-0.0891	-0.104			
Treatment 14	\$162,400	64/68		22	-0.1177	-0.0037	-0.181	-0.0448	0.181			
Control 14				26	0	0.1480	-0.423	-0.2588	-0.169			

Notes: Control N contains taxpayers whose taxable income in year 1 is between \$900 and \$1,600. Control 0 contains all taxpayers below Treatment 1 with taxable income in year 1 abo

Table II: Summary statistics for Single Taxpayers

Groups	Kinks			Number of observations	log(1-T ¹ p/1-T ¹)	log(1-T ² /1-T ¹)	dlog of taxable income	dlog of adjusted gross income	dlog of wages	
	Location	Jump	(1)							(2)
Control N				1,633	0	-0.0911	0.535	0.263	0.272	
Control 0				1,341	0	-0.113	0.316	0.203	0.205	
Treatment 1	\$2,300	0/14		741	-0.1508	-0.120	0.174	0.0687	0.0615	
Control 1				1,764	0	0.0057	0.0959	0.116	0.118	
Treatment 2	\$3,400	14/16		879	-0.0233	-0.00335	0.0264	0.0395	0.0698	
Control 2				1,045	0	0.0056	0.0211	0.0478	0.0376	
Treatment 3	\$4,400	16/18		975	-0.0243	-0.0042	0.0142	0.0106	-0.0048	
Control 3				2,678	0	0.0096	-0.0512	-0.0392	-0.046	
Treatment 4	\$6,500	18/19		1,155	-0.0121	-0.00396	-0.066	-0.072	-0.078	
Control 4				1,770	0	-0.0102	-0.039	-0.0378	-0.0407	
Treatment 5	\$8,500	19/21		1,273	-0.0253	-0.0189	-0.0727	-0.0641	-0.0828	
Control 5				1,619	0	-0.0082	-0.090	-0.0718	-0.0684	
Treatment 6	\$10,800	21/24		1,161	-0.0387	-0.0276	-0.0834	-0.0725	-0.0981	
Control 6				868	0	-0.0031	-0.0926	-0.0778	-0.0884	
Treatment 7	\$12,900	24/26		1,085	-0.0263	-0.0235	-0.0769	-0.0753	-0.0681	
Control 7				522	0	-0.0183	-0.0781	-0.0764	-0.091	
Treatment 8	\$15,000	26/30		972	-0.0555	-0.0293	-0.0974	-0.0866	-0.099	
Control 8				810	0	0.0026	-0.0888	-0.067	-0.084	
Treatment 9	\$18,200	30/34		687	-0.0587	-0.0241	-0.0810	-0.0864	-0.0777	
Control 9				900	0	-0.0016	-0.0753	-0.0560	-0.0896	
Treatment 10	\$23,500	34/39		384	-0.0790	-0.0373	-0.1137	-0.0856	-0.087	
Control 10				234	0	0.0041	-0.093	-0.075	-0.112	
Treatment 11	\$28,800	39/44		177	-0.0856	-0.0515	-0.0946	-0.0633	-0.100	
Control 11				91	0	-0.0170	-0.055	-0.0368	-0.0824	
Treatment 12	\$34,100	44/49		67	-0.0932	-0.0203	-0.210	-0.134	-0.081	
Control 12				47	0	0.0375	-0.171	-0.119	-0.490	
Treatment 13	\$41,500	49/55		25	-0.1256	-0.0370	-0.094	-0.0615	-0.0528	
Control 13				48	0	0.0390	-0.0975	-0.134	-0.108	
Treatment 14	\$55,300	55/63		11	-0.1960	-0.169	0.0271	0.0453	-0.049	
Control 14				25	0	0.0651	-0.142	-0.099	-0.0583	
Treatment 15	\$81,800	63/68		3	-0.1450	0.0652	-0.304	-0.328	-0.254	
Control 15				4	0	0.0565	-0.111	-0.0674	-0.416	

Notes: Control N contains taxpayers whose taxable income in year 1 is between \$900 and \$1,500. Control 0 contains all taxpayers below Treatment 1 with taxable income in year 1 at

Table III: Wald estimates

Kink	Taxable income					AGI	Wages	
	Difference	Difference	Reduced form estimate	Wald estimate	Wald estimate			
Location	jump in marginal rates	treatment/control in dlog(taxinc)	log(1-T _p /1-T ₁) treatment group	treatment/control in log(1-T ₂ /1-T ₁)	((3) divided by (4))	((3) divided by (5))	Wald estimate	
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
PANEL A: Married taxpayers								
\$16,000	21/24	0.0006	-0.0387	-0.0200	-0.016 (0.281)	-0.032 (0.569)	-0.188 (0.398)	0.016 (0.530)
\$20,200	24/28	-0.0110	-0.0545	-0.0255	0.202 (0.173)	0.431 (0.376)	0.403 (0.298)	0.229 (0.371)
\$24,600	28/32	-0.0192	-0.0576	-0.0235	0.333* (0.163)	0.817 (0.483)	0.464 (0.333)	-0.327 (0.377)
\$29,900	32/37	-0.0199	-0.0769	-0.0330	0.258* (0.130)	0.602 (0.340)	0.363 (0.260)	0.838* (0.364)
\$35,200	37/43	-0.0149	-0.0998	-0.0415	0.172 (0.120)	0.398 (0.335)	0.288 (0.281)	0.275 (0.423)
\$45,800	43/49	-0.0465	-0.1109	-0.0488	0.419* (0.162)	0.987* (0.510)	0.580 (0.429)	0.027 (0.554)
\$60,000	49/54	0.0083	-0.1031	-0.0633	-0.081 (0.300)	-0.131 (0.503)	-0.331 (0.455)	0.206 (0.841)
PANEL B: Singles taxpayers								
\$10,800	21/24	0.0075	-0.0387	-0.0212	-0.194 (0.485)	-0.354 (0.782)	-0.071 (0.764)	1.133 (1.103)
\$15,000	26/30	-0.0128	-0.0555	-0.0237	0.230 (0.317)	0.540 (0.832)	0.671 (0.815)	0.489 (0.865)
\$18,200	30/34	0.0007	-0.0587	-0.0245	-0.012 (0.299)	-0.028 (0.669)	1.052 (0.861)	-0.327 (0.659)
\$23,500	34/39	-0.0347	-0.0790	-0.0369	0.440 (0.306)	0.942 (0.711)	0.661 (0.609)	-0.184 (0.598)
\$28,800	39/44	-0.0122	-0.0856	-0.0497	0.143 (0.467)	0.246 (0.850)	-0.019 (0.627)	-0.056 (0.720)

Notes: The numbers in column (3) are calculated, using Tables I and II, as the difference between income in treatment group and the average of the income in the two surrounding control groups (weighted by the number of observations). Similar calculations are performed in column (5) for the change in marginal rates. Standard errors in parenthesis. * for estimates significant at 5% level.

Table IV: 2SLS Elasticity Estimates
All income earners and Middle income earners

	All income earners: Taxable income (1979 \$)			Middle income earners: Taxable income (1979 \$)		
	Singles: \$3,000-\$40,000-Married: \$5,000-\$70,000			Singles: \$12,000-\$28,000. Married: \$16,000-\$36,000		
	Taxable income	AGI	Wages	Taxable income	AGI	Wages
	(1)	(2)	(3)	(4)	(5)	(6)
PANEL A: Married and Single taxpayers						
PANEL A1: Itemizers and non itemizers						
log(1-T ²)/(1-T ¹)	0.282 (0.199)	0.181 (0.157)	0.080 (0.188)	0.395* (0.199)	0.334* (0.165)	0.120 (0.196)
N. obs.	49,816	50,326	44,993	21,018	21,084	19,800
PANEL A2: Itemizers						
log(1-T ²)/(1-T ¹)	0.393 (0.244)	0.356* (0.178)	0.105 (0.232)	0.619* (0.265)	0.374* (0.197)	0.096 (0.246)
N. obs.	18,764	18,906	17,210	11,546	11,590	11,003
PANEL A3: Non itemizers						
log(1-T ²)/(1-T ¹)	-0.046 (0.296)	-0.089 (0.241)	-0.012 (0.287)	0.017 (0.298)	0.191 (0.271)	0.183 (0.320)
N. obs.	31,052	31,420	27,783	9,472	9,494	8,797
PANEL B: Married taxpayers						
PANEL B1: Itemizers and non itemizers						
log(1-T ²)/(1-T ¹)	0.389 (0.217)	0.202 (0.154)	0.087 (0.197)	0.437* (0.240)	0.383* (0.190)	0.272 (0.242)
N. obs.	30,675	30,929	28,260	15,630	15,675	14,947
PANEL B2: Itemizers						
log(1-T ²)/(1-T ¹)	0.651* (0.274)	0.421* (0.186)	0.231 (0.244)	0.705* (0.305)	0.521* (0.234)	0.332 (0.289)
N. obs.	15,924	16,033	15,015	9,964	9,998	9,632
PANEL B3: Non itemizers						
log(1-T ²)/(1-T ¹)	-0.091 (0.345)	-0.193 (0.252)	-0.167 (0.327)	-0.148 (0.384)	0.028 (0.314)	0.114 (0.436)
N. obs.	14,751	14,896	13,245	5,666	5,677	5,315
PANEL C: Single taxpayers						
log(1-T ²)/(1-T ¹)	0.170 (0.451)	0.188 (0.376)	-0.077 (0.406)	0.275 (0.442)	0.472 (0.454)	-0.155 (0.439)
N. obs.	19,141	19,397	16,733	5,388	5,409	4,853

Notes: All regressions include $\log(z_1)$, taxinc^1 , taxinc^2 , taxinc^3 and taxinc^4 as control variables. Regressions in panel A control in addition for marital status. Regressions including both itemizers and non itemizers control in addition for itemizer status. Standard errors in parenthesis. * for estimates significant at 5% level.

Table V: 2SLS Elasticity Estimates
High income earners and Low income earners

	High income taxpayers			Low income taxpayers	
	Taxable income (1979 dollars)			Taxable income (1979 dollars)	
	Singles:\$21,000-\$65,000-Married:\$31,000-\$90,000			Singles:\$0-\$3,400-Married:\$0-\$5,000	
	Dependent variable			Dependent variable	
	Taxable income	Adjusted gross income	Wages	Adjusted gross income	Wages
	(1)	(2)	(3)	(4)	(5)
PANEL A: Married and Single taxpayers					
log(1-T ²)/(1-T ¹)	0.277 (0.252)	0.022 (0.197)	-0.441 (0.282)		
N. obs.	4,618	4,629	4,174		
PANEL B: Married taxpayers					
log(1-T ²)/(1-T ¹)	0.332 (0.268)	0.067 (0.218)	-0.342 (0.335)	-0.289 (0.281)	-0.052 (0.457)
N. obs.	3,466	3,474	3,207	3,895	2,733
PANEL C: Single taxpayers					
log(1-T ²)/(1-T ¹)	0.159 (0.597)	-0.223 (0.409)	-0.587 (0.495)	1.082* (0.433)	1.310* (0.480)
N. obs.	1,152	1,155	967	8,713	7,622

Notes: Regressions for high incomes include log(z1), taxinc1, taxinc1², taxinc1³ and taxinc1⁴ as controls. Regressions for low incomes include log(z1), taxinc1 as control variables. All regressions include itemization status as control variables. Regressions in panel A control in addition for marital status. Standard errors in parenthesis. * for estimates significant at 5% level.

Bibliography

- [1] Angrist, Joshua and Victor Lavy. "Using Maimonides Rule to Estimate the Effect of Class Size on Scholastic Achievement." forthcoming *Quarterly Journal of Economics*, 1999.
- [2] Atkinson, Anthony B. "Public Economics and the Economic Public." *European Economic Review*, 1990, 34(2-3), pp. 225-248.
- [3] Atkinson, Anthony B. "Maxi-Min and Optimal Income Taxation." *Cahiers du Séminaire d'Econométrie*, 1975, pp. 73-86.
- [4] Atkinson, Anthony and Joseph Stiglitz. *Lectures on Public Economics*, London: McGraw-Hill, 1980.
- [5] Auerbach, Alan J. "The Theory of Excess Burden and Optimal Taxation." in A.J. Auerbach and M. Feldstein (eds.), *Handbook of Public Economics*, Amsterdam: North-Holland, 1985.
- [6] Auerbach, Alan and Joel Slemrod. "The Economic Effects of the Tax Reform Act of 1986." *Journal of Economic Literature*, 1997, 35(2), pp. 589-632.
- [7] Auten, Gerald and Robert Carroll. "The Effect of Income Taxes on Household Behavior." Office of Tax Analysis, U.S. Department of the Treasury, mimeo 1997, forthcoming *Review of Economics and Statistics*.
- [8] Ballard, Charles L. and Don Fullerton. "Distortionary Taxes and the Provision of Public Goods." *Journal of Economic Perspectives*, Summer 1992, 6(3), pp. 117-31.

- [9] Baron, David P. and Roger B. Myerson. "Regulating a Monopolist with Unknown Costs." *Econometrica*, 1982, 50, pp. 911-30.
- [10] Boiteux, Marcel. "Sur la gestion des monopoles publics astreints à l'équilibre budgétaire." *Econometrica*, 1956, 24, pp. 22-40. Translated as "On the Management of Public Monopolies Subject to Budgetary Constraints." *Journal of Economic Theory*, 1971, 3, pp. 219-240.
- [11] Bound, John and George Johnson. "Changes in the Structure of Wages in the 1980s: An Evaluation of Alternative Explanations." *American Economic Review*, 82, 1992, pp. 371-392.
- [12] Chipman, John S. and James C. Mcore. "Compensating Variation, Consumer's Surplus and Welfare." *American Economic Review*, 70(5), 1980, pp. 933-49.
- [13] Christian, Charles and Peter Frischmann. "Attrition in the Statistics of Income Panel of Individual Returns." *National Tax Journal*, 1989, 62(4), pp. 495-501.
- [14] Dahlby, Bev. "Progressive Taxation and the Social Marginal Cost of Public Funds." *Journal of Public Economics*, 1998, 67(1), pp. 105-122.
- [15] 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.
- [16] Ebert, Udo. "A Reexamination of the Optimal Nonlinear Income Tax." *Journal of Public Economics*, 1992, 49(1), pp. 47-73.
- [17] Eissa, Nada. "Tax Reforms and Labor Supply." in *Tax Policy and the Economy*, vol. 10, ed. James Poterba, Cambridge:MIT Press, pp.119-51, 1996.
- [18] Feenberg, Daniel and James Poterba. "Income Inequality and the Incomes of Very High-Income Taxpayers: Evidence from Tax Returns." in *Tax Policy and the Economy*, vol. 7, ed. James Poterba. Cambridge: MIT Press, 1993.
- [19] Feldstein, Martin. "The Effect of Marginal Tax Rates on Taxable Income: A Panel Study of the 1986 Tax Reform Act." *Journal of Political Economy*, 1995, 103(3), pp. 551-572.

- [20] Freeman, Richard. "How Much has De-Unionization Contributed to the Rise in Male Earnings Inequality." in *Uneven Tides: Rising Inequality in America*, ed. P. Gottschalk and S. Danziger. New-York: Russell Sage Foundation, 1993.
- [21] Gibson, John. *The Politics and Economics of the Poll-Tax: Mrs Thatcher's Downfall*, EMAS Ltd, United Kingdom, 1990.
- [22] Goolsbee, Austan. "What Happens When You Tax the Rich? Evidence from Executive Compensation.", NBER Working Paper, No. 6333, 1997.
- [23] Gordon, Roger and Joel Slemrod. "Are 'Real' Responses to Taxes Simply Income Shifting Between Corporate and Personal Tax Bases?" in *Does Atlas Shrug? The Economic Consequences of Taxing the Rich*, ed. J. Slemrod, Cambridge University Press, forthcoming 1999.
- [24] Harberger, Arnold C. "Principles of Efficiency: The Measurement of Waste." *American Economic Review*, 1964, 54(3), pp. 58-76.
- [25] Harberger, Arnold C. "On the Use of Distributional Weights in Social Cost-Benefit Analysis.", *Journal of Political Economics*, 1978, 86(2), pp. S87-S120.
- [26] Hausman, Jerry. "Labor Supply." in *How Taxes Affect Economic Behavior*, ed. H. Aaron and J. Pechman. Washington, D.C.: Brookings Institution, 1981.
- [27] Hausman, Jerry, "Taxes and Labor Supply." in A.J. Auerbach and M. Feldstein (eds.), *Handbook of Public Economics*, Amsterdam: North-Holland, 1985.
- [28] Heckman, James. "What has been learned about Labor Supply in the Past Twenty Years?", *American Economic Review*, 1993, 83(2), pp. 116-121.
- [29] Katz, Lawrence and Kevin Murphy. "Changes in Relative Wages, 1963-1987: Supply and Demand Factors." *Quarterly Journal of Economics*, 1992, 107, pp. 35-78.
- [30] Kay, J.A. "The Deadweight Loss from a Tax System." *Journal of Public Economics*, 1980, (13), pp. 111-119.

- [31] Levy, Frank and Richard Murnane. "U.S. Earnings Levels and Earnings Inequality: A Review of Recent Trends and Proposed Explanations." *Journal of Economic Literature*, 1992, 30, pp. 1333-81.
- [32] Lindsey, Lawrence. "Individual Taxpayer Response to Tax Cuts: 1982-1984, with Implications for the Revenue Maximizing Tax Rate." *Journal of Public Economics*, 1987, (33), pp. 173-206.
- [33] MaCurdy, Thomas, David Green and Harry Paarsch. "Assessing Empirical Approaches for Analyzing Taxes and Labor Supply." *The Journal of Human Resources*, Fall 1990, pp. 415-490.
- [34] Mayshar, Joram. "On Measuring the Marginal Cost of Funds Analytically." *American Economic Review*, 1991, 81(5), pp. 1329-35.
- [35] Mirrlees, James A. "An Exploration in the Theory of Optimal Income Taxation." *Review of Economic studies*, 1971, 38, pp. 175-208.
- [36] Mirrlees, James A. "Optimal Tax Theory: a Synthesis." *Journal of Public Economics*, 1976, 6, pp. 327-58.
- [37] Mirrlees, James A. "The Theory of Optimal Taxation." in K.J. Arrow and M.D. Intrilligator (eds.), *Handbook of Mathematical Economics*, Amsterdam: North-Holland, 1986.
- [38] Mroz, Thomas. "The Sensitivity of an Empirical Model of Married Women's Hours of Work to Economic and Statistical Assumptions." *Econometrica*, 55, 1987 pp. 765-799.
- [39] Murphy, Kevin and Finis Welch. "The Structure of Wages." *Quarterly Journal of Economics*, 1992, 107(1), pp. 285-326.
- [40] Mussa, Michael and Sherwin Rosen. "Monopoly and Product Quality." *Journal of Economic Theory*, 1978, 28, pp. 265-281.
- [41] Myles, Gareth D. *Public Economics*. Cambridge: Cambridge University Press, 1995.

- [42] Navratil, John. "The Economic Recovery Tax Act of 1981: Evidence on Individual Taxpayer Behavior from Panel Tax Return Data." Unpublished Harvard Thesis 1995.
- [43] Pencavel, John. "Labor Supply of Men." in O. Ashenfelter and R. Layard (eds.), *Handbook of Labor Economics*, 1986, Amsterdam: North-Holland, pp. 3-102.
- [44] Piketty, Thomas. "La Redistribution Fiscale face au Chômage." *Revue Française d'Economie*, 1997, 12(1), pp. 157-201.
- [45] Ramsey, Frank. "A Contribution to the Theory of Taxation." *Economic Journal*, 1927, 37, pp. 47-61.
- [46] Sadka, Efraim. "On Income Distribution, Incentive Effects and Optimal Income Taxation." *Review of Economic Studies*, 1976, 43(2), pp. 261-8.
- [47] Samuelson, Paul A. "Principles of Efficiency-Discussion." *American Economic Review*, 1964, 54(3), pp. 93-96.
- [48] Seade, Jesus K. "On the Shape of Optimal Tax Schedules." *Journal of Public Economics*, 1977, 7(2), pp. 203-236.
- [49] Seade, Jesus K. "On the Sign of the Optimum Marginal Income Tax." *Review of Economic Studies*, 1982, 49(4), pp. 637-643.
- [50] Sheshinski, Eytan. "The Optimal Linear Income Tax." *Review of Economic Studies*, 1972, 39, pp. 297-302.
- [51] Slemrod, Joel. "Taxation and Inequality: A Time-exposure Perspective." *Tax Policy and the Economy*, 1992, pp. 105-127.
- [52] Slemrod, Joel. "Income Creation or Income Shifting? Behavioral Responses to the Tax Reform Act of 1986." *American Economic Review*, 1995, 85(2), pp. 175-80.
- [53] Slemrod, Joel. "High Income Families and the Tax Changes of the 1980s", in *Empirical Foundations of Household Taxation*, eds. M. Feldstein and J. Poterba, University of Chicago, 1996.

- [54] Stern, N. H. "On the Specification of Models of Optimal Taxation." *Journal of Public Economics*, 1976, 6, pp. 123-62.
- [55] Steuerle, C. Eugene. *The Tax Decade: How Taxes Came to Dominate the Public Agenda*. The Urban Institute Press, Washington, D.C., 1991.
- [56] Stiglitz, Joseph E. "Pareto Efficient and Optimal Taxation and the New Welfare Economics." in A.J. Auerbach and M. Feldstein (eds.), *Handbook of Public Economics*, Amsterdam: North-Holland, Volume II, 1987, pp. 991-1042 .
- [57] Topel, Robert. "Regional Labor Markets and the Determinants of Wages Inequality." *American Economic Review*, 1994, 84, pp. 17-22.
- [58] Triest, R. "The Effect of Income Taxation on Labor Supply in the United-States." *Journal of Human Resources*, 1990, 25, (Summer), pp. 491-516.
- [59] Tuomala, Matti. *Optimal Income Tax and Redistribution*. Oxford: Clarendon Press, 1990.
- [60] Van der Klaauw, W. "A Regression-Discontinuity Evaluation of the Effect of Financial Aid Offers on College Enrollment." unpublished mimeograph, New York University, 1996.
- [61] Wilson, Robert, B. *Nonlinear Pricing*, Oxford: Oxford University Press, 1993.