Optimal Labor Income Taxation

Thomas Piketty[*] and Emmanuel Saez[†,‡]

[*]Paris School of Economics, Paris, France
[†]Department of Economics, University of California, 530 Evans Hall #3880, Berkeley, CA 94720, USA
[‡]National Bureau of Economic Research, USA

Contents

1. INTRODUCTION

This handbook chapter considers optimal labor income taxation, that is, the fair and efficient distribution of the tax burden across individuals with different earnings. A large academic literature has developed models of optimal tax theory to cast light on this issue. Models in optimal tax theory typically posit that the tax system should maximize a social welfare function subject to a government budget constraint, taking into account how individuals respond to taxes and transfers. Social welfare is larger when resources are more equally distributed, but redistributive taxes and transfers can negatively affect incentives to work and earn income in the first place. This creates the classical trade-off between equity and efficiency which is at the core of the optimal labor income tax problem.

In this chapter, we present recent developments in the theory of optimal labor income taxation. We emphasize connections between theory and empirical work that were previously largely absent from the optimal income tax literature. Therefore, throughout the chapter, we focus less on formal modeling and rigorous derivations than was done in previous surveys on this topic (Atkinson & Stiglitz, 1980; Kaplow, 2008; Mirrlees (1976, 1986, chap. 24); Stiglitz, 1987, chap. 15; Tuomala, 1990) and we try to systematically connect the theory to both real policy debates and empirical work on behavioral responses to taxation.[1] This chapter limits itself to the analysis of optimal *labor income* taxation and related means-tested transfers.[2]

First, we provide historical and international background on labor income taxation and transfers. In our view, knowing actual tax systems and understanding their history and the key policy debates driving their evolution is critical to guide theoretical modeling and successfully capture the first order aspects of the optimal tax problem. We also briefly review the history of the field of optimal labor income taxation to place our chapter in its academic context.

Second, we review the theoretical underpinnings of the standard optimal income tax approach, such as the social welfare function, the fallacy of the second welfare theorem, and hence the necessity of tackling the equity-efficiency trade off. We also present the

---

[1] Boadway (2012) also provides a recent, longer, and broader survey that aims at connecting theory to practice.

[2] The analysis of optimal *capital income* taxation naturally involves dynamic considerations and is covered in the chapter by Kopczuk in this volume.

key parameters capturing labor supply responses as they determine the efficiency costs of taxation and hence play a crucial role in optimal tax formulas.

Third, we present the simple model of optimal linear taxation. Considering linear labor income taxation simplifies considerably the exposition but still captures the key equity-efficiency trade-off. The derivation and the formula for the optimal linear tax rate are also closely related to the more complex nonlinear case, showing the tight connection between the two problems. The linear tax model also allows us to consider extensions such as tax avoidance and income shifting, random earnings, and median voter tax equilibria in a simpler way.

Fourth, we consider optimal nonlinear income taxation with particular emphasis on the optimal top tax rate and the optimal profile of means-tested transfers at the bottom. We consider several extensions including extensive labor supply responses, international migration, or rent-seeking models where pay differs from productivity.

Fifth, we consider additional deeper extensions of the standard model including tagging (i.e., conditioning taxes and transfers on characteristics correlated with ability to earn), the use of differential commodity taxation to supplement the income tax, the use of in-kind transfers (instead of cash transfers), the treatment of couples and children in tax and transfer systems, or models with relative income concerns. Many of those extensions cannot be satisfactorily treated within the standard utilitarian social welfare approach. Hence, in a number of cases, we present the issues only heuristically and leave formal full-fledged modeling to future research.

Sixth and finally, we come back to the limitations of the standard utilitarian approach that have appeared throughout the chapter. We briefly review the most promising alternatives. While many recent contributions use general Pareto weights to avoid the strong assumptions of the standard utilitarian approach, the Pareto weight approach is too general to deliver practical policy prescriptions in most cases. Hence, it is important to make progress both on normative theories of justice stating how social welfare weights should be set and on positive analysis of how individual views and beliefs about redistribution are formed.

Methodologically, a central goal of optimal tax analysis should be to cast light on actual tax policy issues and help design better tax systems. Theory and technical derivations are very valuable to rigorously model the problem at hand. A key aim of this chapter is to show how to make such theoretical findings applicable. As argued in Diamond and Saez (2011), theoretical results in optimal tax analysis are most useful for policy recommendations when three conditions are met. (1) Results should be based on economic mechanisms that are empirically relevant and first order to the problem at hand. (2) Results should be reasonably robust to modeling assumptions and in particular to the presence of heterogeneity in individual preferences. (3) The tax policy prescription needs to be implementable—that is, the tax policy needs to be relatively easy to explain

and discuss publicly, and not too complex to administer relative to actual practice.[3] Those conditions lead us to adopt two methodological choices.

First, we use the "sufficient statistics" approach whereby optimal tax formulas are derived and expressed in terms of estimable statistics including social marginal welfare weights capturing society's value for redistribution and labor supply elasticities capturing the efficiency costs of taxation (see Chetty, 2009a for a recent survey of the "sufficient statistics" approach in public economics). This approach allows us to understand the key economic mechanisms behind the formulas, helping meet condition (1). The "sufficient statistics" formulas are also often robust to change the primitives of the model, which satisfies condition (2).

Second, we tend to focus on simple tax structures—e.g., a linear income tax—without systematically trying to derive the most general tax system possible. This helps meet condition (3) as the tax structures we obtain will by definition be within the realm of existing tax structures.[4] This is in contrast to the "mechanism design" approach that derives the most general optimum tax compatible with the informational structure. This "mechanism design" approach tends to generate tax structures that are highly complex and results that are sensitive to the exact primitives of the model. The mechanism design approach has received renewed interest in the new dynamic public finance literature that focuses primarily on dynamic aspects of taxation.[5]

The chapter is organized as follows. Section 2 provides historical and international background on labor income taxation and means-tested transfers, and a short review of the field of optimal labor income taxation. Section 3 presents the key concepts: the standard utilitarian social welfare approach, the fallacy of the second welfare theorem, and the key labor supply concepts. Section 4 discusses the optimal linear income tax problem. Section 5 presents the optimal nonlinear income taxation problem with particular emphasis on the optimal top tax rate and the optimal profile of means-tested transfers. Section 6 considers a number of extensions. Section 7 discusses limits of the standard utilitarian approach.

2. BACKGROUND ON ACTUAL TAX SYSTEMS AND OPTIMAL TAX THEORY

2.1. Actual Tax Systems

Taxes. Most advanced economies in the OECD raise between 35% and 50% of national income (GNP net of capital depreciation) in taxes. As a first approximation, the share

---

[3] Naturally, the set of possible tax systems evolves overtime with technological progress. If more complex tax innovations become feasible and can realistically generate large welfare gains, they are certainly worth considering.

[4] The simple tax structure approach also helps with conditions (1) and (2) as the economic trade-offs are simpler and more transparent, and the formulas for simple tax structures tend to easily generalize to heterogeneous populations.

[5] See Golosov, Tsyvinski, and Werning (2006) and Kocherlakota (2010) for recent surveys of the new dynamic public finance literature. Piketty and Saez (2012a,b) analyze the problem optimal taxation of capital and inheritances in a dynamic model but using a sufficient statistics approach and focusing on simple tax structures.

of total tax burden falling on capital income roughly corresponds to the share of capital income in national income (i.e., about 25%).[6] The remaining 75% of taxes falls on labor income (OECD 2011a),[7] which is the part we are concerned with in this chapter.

Historically, the overall tax to national income ratio has increased substantially during the first part of the 20th century in OECD countries from about 10% on average around 1900 to around 40% by 1970 (see e.g., Flora, 1983 for long time series up to 1975 for a number of Western European countries and OECD, *Revenue Statistics*, OECD, 2011a for statistics since 1965). Since the late 1970s, the tax burden in OECD countries has been roughly stable. The share of taxes falling on capital income has declined slightly in Europe and has been approximately stable in the United States.[8] Similar to the historical evolution, tax revenue to national income ratios increase with GDP per capita when looking at the current cross-section of countries. Tax to national income ratios are smaller in less developed and developing countries and higher on average among the most advanced economies.

To a first approximation, the tax burden is distributed proportionally to income. Indeed, the historical rise in the tax burden has been made possible by the ability of the government to monitor income flows in the modern economy and hence impose payroll taxes, profits taxes, income taxes, and value-added-taxes, based on the corresponding income and consumption flows. Before the 20th century, the government was largely limited to property and presumptive taxes, and taxes on a few specific goods for which transactions were observable. Such archaic taxes severely limited the tax capacity of the government and tax to national income ratios were low (see Ardant, 1971 and Webber & Wildavsky, 1986 for a detailed history of taxation). The transition from archaic to broad-based taxes involves complex political and administrative processes and may occur at various speeds in different countries.[9]

In general, actual tax systems achieve some tax progressivity, i.e., tax rates rising with income, through the individual income tax. Most individual income tax systems have brackets with increasing marginal tax rates. In contrast, payroll taxes or consumption taxes tend to have flat rates. Most OECD countries had very progressive individual income

---

[6] This is defining taxes on capital as the sum of property and wealth taxes, inheritance and gift taxes, taxes of corporate and business profits, individual income taxes on individual capital income, and the share of consumption taxes falling on capital income. Naturally, there are important variations over time and across countries in the relative importance of these various capital tax instruments. See e.g., Piketty and Saez (2012a).

[7] Including payroll taxes, individual income tax on labor income, and the share of consumption taxes falling on labor income.

[8] Again, there are important variations in capital taxes which fall beyond the scope of this chapter. In particular, corporate tax rates have declined significantly in Europe since the early 1990s (due to tax competition), but tax revenues have dropped only slightly, due to a global rise in the capital share, the causes of which are still debated. See e.g., Eurostat (2012).

[9] See e.g., Piketty and Qian (2009) for a contrast between China (where the income tax is about to become a mass tax, like in developed countries) and India (where the income tax is still very much an elite tax raising limited revenue). Cagé and Gadenne (2012) provide a comprehensive empirical analysis of the extent to which low- and middle-income countries were able to replace declining trade tax revenues by modern broad based taxes since the 1970s. See Kleven, Kreiner and Saez (2009b) for a theoretical model of the fiscal modernization process.
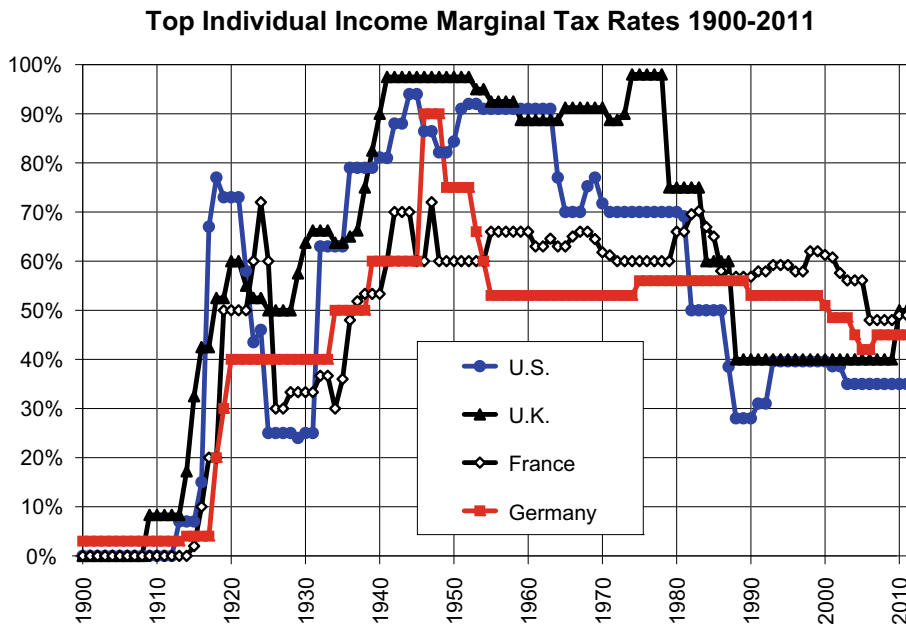
Top Individual Income Marginal Tax Rates 1900-2011

Figure 1 Top Marginal income tax rates in the US, UK, France, Germany. This figure, taken from Piketty et al. (2011), depicts the top marginal individual income tax rate in the US, UK, France, Germany since 1900. The tax rate includes only the top statutory individual income tax rate applying to ordinary income with no tax preference. State income taxes are not included in the case of the United States. For France, we include both the progressive individual income tax and the flat rate tax "Contribution Sociale Généralisée."

taxes in the post–World War II decades with a large number of tax brackets and high top tax rates (see e.g., OECD, 1986). Figure 1 depicts top marginal income tax rate in the United States, the United Kingdom, France, and Germany since 1900. When progressive income taxes were instituted—around 1900–1920 in most developed countries, top rates were very small—typically less than 10%. They rose very sharply in the 1920–1940s, particularly in the US and in the UK. Since the late 1970s, top tax rates on upper income earners have declined significantly in many OECD countries, again particularly in English speaking countries. For example, the US top marginal federal individual tax rate stood at an astonishingly high 91% in the 1950–1960s but is only 35% today (Figure 1). Progressivity at the very top is often counter balanced by the fact that a substantial fraction of capital income receives preferential tax treatment under most income tax rules.[10]

---

[10] For example, (Landais, Piketty, and Saez (2011)) show that tax rates decline at the very top of the French income distribution because of such preferential tax treatment and of various tax loopholes and fiscal optimization strategies. In the United States as well, income tax rates decline at the very top due to the preferential treatment of realized capital gains which constitute a large fraction of top incomes (US Treasury, 2012). See Piketty and Saez (2007) for an analysis of progressivity of the federal tax system since 1960. Note that preferential treatment for capital income did not exist when modern income taxes were created in 1900–1920. Preferential treatment was developed mostly in the

Table 1 Public Spending in OECD Countries (2000–2010, Percent of GDP)

	US (1)	Germany (2)	France (3)	UK (4)	Total OECD (5)
Total public spending	**35.4%**	**44.1%**	**51.0%**	**42.1%**	**38.7%**
Social public spending	**22.4%**	**30.6%**	**34.3%**	**26.2%**	**25.1%**
Education	4.7%	4.4%	5.2%	4.8%	4.9%
Health	7.7%	7.8%	7.1%	6.1%	5.6%
Pensions	6.0%	10.1%	12.2%	4.8%	6.5%
Income support to working age	2.7%	3.9%	4.8%	4.9%	4.4%
Other social public spending	1.3%	4.4%	5.1%	5.7%	3.7%
Other public spending	**13.0%**	**13.5%**	**16.7%**	**15.9%**	**13.6%**

Notes and sources: OECD Economic Outlook 2012, Annex Tables 25–31; Adema et al., 2011, Table 1.2; Education at a Glance, OECD 2011, Table B4.1. Total public spending includes all government outlays (except net debt interest payments). Other social public spending includes social services to the elderly and the disabled, family services, housing and other social policy areas (see Adema et al., 2011, p.21). We report 2000–2010 averages so as to smooth business cycle variations. Note that tax to GDP ratios are a little bit lower than spending to GDP ratios for two reasons: (a) governments typically run budget deficits (which can be large, around 5–8 GDP points during recessions), (b) governments get revenue from non-tax sources (such as user fees, profits from government owned firms, etc.).

As we shall see, optimal nonlinear labor income tax theory derives a simple formula for the optimal tax rate at the top of the earnings distribution. We will not deal however with the dynamic redistributive impact of tax progressivity through capital and wealth taxation, which might well have been larger historically than its static impact, as suggested by the recent literature on the long run evolution of top income shares.[11]

Transfers. The secular rise in taxes has been used primarily to fund growing public goods and social transfers in four broad areas: education, health care, retirement and disability, and income security (see Table 1). Indeed, aside from those four areas, government spending (as a fraction of GDP) has not grown substantially since 1900. All advanced economies provide free public education at the primary and secondary level, and heavily subsidized (and often almost free) higher education.[12] All advanced economies except the United States provide universal public health care (the United States provides public health care

postwar period in order to favor savings and reconstruction, and then extended since the 1980–1990s in the context of financial globalization and tax competition. For a detailed history in the case of France, see Piketty (2001).

[11] See Atkinson, Piketty and Saez (2011) for a recent survey. One of the main findings of this literature is that the historical decline in top income shares that occurred in most countries during the first half of the twentieth century has little to do with a Kuznets-type process. It was largely due to the fall of top capital incomes, which apparently never fully recovered from the 1914–1945 shocks, possibly because of the rise of progressive income and estate taxes and their dynamic impact of savings, capital accumulation and wealth concentration.

[12] Family benefits can also be considered as part of education spending. Note that the boundaries between the various social spending categories reported on Table 1 are not entirely homogenous across OECD countries (e.g., family benefits are split between "Income support to the working age" and "Other social public spending"). Also differences

to the old and the poor through the Medicare and Medicaid programs respectively, which taken together happen to be more expensive than most universal health care systems), as well as public retirement and disability benefits. Income security programs include unemployment benefits, as well as an array of means-tested transfers (both cash and in-kind). They are a relatively small fraction of total transfers (typically less than 5% of GDP, out of a total around 25–35% of GDP for social spending as a whole; see Table 1).

Education, family benefits, and health care government spending are approximately a demogrant, that is, a transfer of equal value for all individuals in expectation over a life-time.[13] In contrast, retirement benefits are approximately proportional to lifetime labor income in most countries.[14] Finally, income security programs are targeted to lower income individuals. This is therefore the most redistributive component of the transfer system. Income security programs often take the form of in-kind benefits such as subsidized housing, subsidized food purchases (e.g., food stamps and free lunches at school in the United States), or subsidized health care (e.g., Medicaid in the United States). They are also often targeted to special groups such as the unemployed (unemployment insurance), the elderly or disabled with no resources (for example Supplemental Security Income in the United States). Means-tested cash transfer programs for "able bodied" individuals are only a small fraction of total transfers. To a large extent, the rise of the modern welfare state is the rise of universal access to "basic goods" (education, health, retirement and social insurance), and not the rise of cash transfers (see e.g., Lindert, 2004).[15]

In recent years, traditional means-tested cash welfare programs have been partly replaced by in-work benefits. The shift has been particularly large in the United States and the United Kingdom. Traditional means-tested programs are L-shaped with income. They provide the largest benefits to those with no income and those benefits are then phased-out at high rates for those with low earnings. Such a structure concentrates benefits among those who need them most. At the same time and as we shall see, these phase-outs discourage work as they create large implicit taxes for low earners. In contrast, in-work benefits are inversely U-shaped, first rising and then declining with earnings. Benefits are nil for those with no earnings and concentrated among low earners before being phased-out. Such a structure encourages work but fails to provide support to those with no earnings, arguably those most in need of support.

---

in tax treatment of transfers further complicate cross country comparisons. Here we simply care about the broad orders of magnitude. For a detailed cross-country analysis, see Adema, Fron, and Ladaique (2011).

[13] Naturally, higher income individuals are often better able to navigate the public education and health care systems and hence tend to get a better value out of those benefits than lower income individuals. However, the value of those benefits certainly grows less than proportionally to income.

[14] In most countries, benefits are proportional to payroll tax contributions. Some countries—such as the United Kingdom—provide a minimum pension that is closer to a demogrant.

[15] It should be noted that the motivation behind the historical rise of these public services has to do not only with redistributive objectives, but also with the perceived failure of competitive markets in these areas (e.g., regarding the provision of health insurance or education). We discuss issues of individual and market failures in Section 6 below.

Overall, all transfers taken together are fairly close to a demogrant, i.e., are about constant with income. Hence, the optimal linear tax model with a demogrant is a reasonable first order approximation of actual tax systems and is useful to understand how the level of taxes and transfers should be set. At a finer level, there is variation in the profile of transfers. Such a profile can be analyzed using the more complex nonlinear optimal tax models.

Budget Set. The budget set relating pre-tax and pre-transfers earnings to post-tax post-transfer disposable income summarizes the net impact of the tax and transfer system.
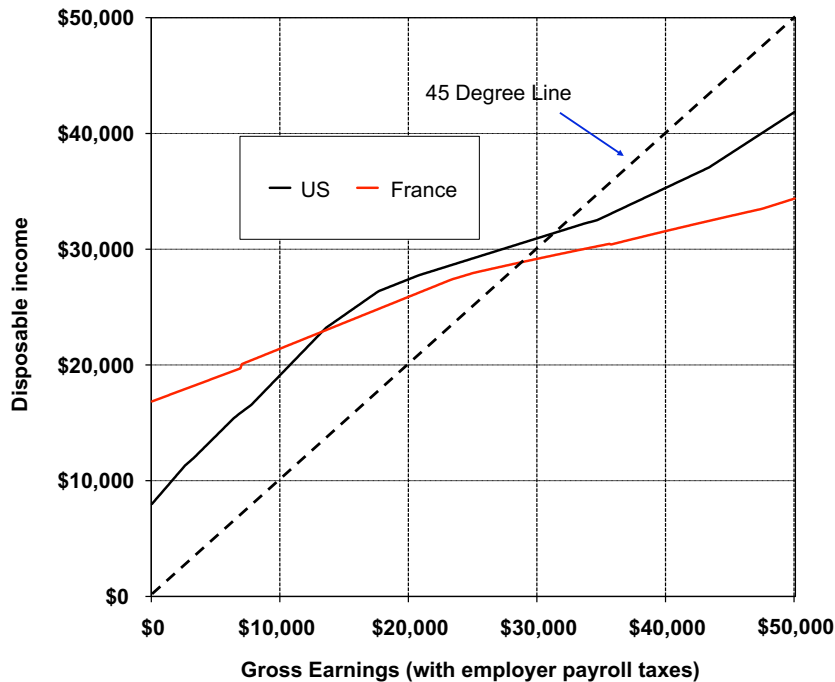

Figure 2 Tax/transfer system in the US and France, 2010, single parent with two children. The figure depicts the budget set for a single parent with two children in France and the United States (exchange rate 1 Euro = $1.3). The figure includes payroll taxes and income taxes on the tax side. It includes means-tested transfer programs (TANF and Food stamps in the United States, and the minimum income–RSA for France) and tax credits (the Earned Income Tax Credit and the Child Tax Credit in the United States, in-work benefit Prime pour l'Emploi and cash family benefits in France). Note that this graph ignores important elements. First, the health insurance Medicaid program in the United States is means tested and adds a significant layer of implicit taxation on low income work. France offers universal health insurance which does not create any additional implicit tax on work. Second, the graph ignores in-kind benefits for children such as subsidized child care and free pre-school kindergarten in France that have significant value for working single parents. Such programs barely exist in the United States. Third, the graph ignores temporary unemployment insurance benefits which depend on previous earnings for those who have become recently unemployed and which are significantly more generous in France both in level and duration.

The slope of the budget set captures the marginal incentive to work. Figure 2 depicts the budget set for a single parent with two children in France and the United States. The figure includes all payroll taxes and the income tax, on the tax side. It includes means-tested transfer programs (TANF and Food Stamps in the United States, and the minimum income—RSA for France) and tax credits (the Earned Income Tax Credit and the Child Tax Credit in the United States, in-work benefit Prime pour l'Emploi and cash family benefits in France). France offers more generous support to single parents with no earnings but the French tax and transfer system imposes higher implicit taxes on work.[16] As mentioned above, optimal nonlinear income tax theory precisely tries to assess what is the most desirable profile for taxes and transfers.

Policy Debate. At the center of the political debate on labor income taxation and transfers is the equity-efficiency trade off. The key argument in favor of redistribution through progressive taxation and generous transfers is that social justice requires the most successful to contribute to the economic well being of the less fortunate. The reasons why society values such redistribution from high to low incomes are many. As we shall see, the standard utilitarian approach posits that marginal utility of consumption decreases with income so that a more equal distribution generates higher social welfare. Another and perhaps more realistic reason is that differences in earnings arise not only from differences in work behavior (over which individuals have control) but also from differences in innate ability or family background or sheer luck (over which individuals have little control). The key argument against redistribution through taxes and transfers is efficiency. Taxing the rich to fund means-tested programs for the poor reduces the incentives to work both among the rich and among transfer recipients. In the standard optimal tax theory, such responses to taxes and transfers are costly solely because of their effect on government finances.

Do Economists Matter? The academic literature in economics does play a role, although often an indirect one, in shaping the debate on tax and transfer policy. In the 1900–1910s, when modern progressive income taxes were created, economists appear to have played a role, albeit a modest one. Utilitarian economists like Jevons, Edgeworth, and Marshall had long argued that the principles of marginal utility and equal sacrifice push in favor of progressive tax rates (see e.g., Edgeworth, 1897)—but such theoretical results had little impact on the public debate. Applied economists like Seligman wrote widely translated and read

---

[16] Note that this graph ignores important elements. First, the health insurance Medicaid program in the United States is means-tested and adds a significant layer of implicit taxation on low income work. France offers universal health insurance which does not create any additional implicit tax on work. Second, the graph ignores in-kind benefits for children such as subsidized child care and free pre-school kindergarten in France that have significant value for working single parents. Such programs barely exist in the United States. Third, the graph ignores housing benefits, which are substantial in France. Fourth, the graph ignores temporary unemployment insurance benefits which depend on previous earnings for those who have become recently unemployed and which are significantly more generous in France both in level and duration. Finally, this graph ignores consumption taxes, implying that the cutoff income level below which transfers exceed taxes is significantly overestimated. This cutoff also greatly varies with the family structure (e.g., able bodied single individuals with no dependent receive zero cash transfers in the US but significant transfers in France).

books and reports (see e.g., Seligman, 1911) arguing that progressive income taxation was not only fair but also economically efficient and administratively manageable.[17] Such arguments expressed in terms of practical economic and administrative rationality helped to convince reluctant mainstream economists in many countries that progressive income taxation was worth considering.[18]

In the 1920–1940s, the rise of top tax rates seems to have been the product of public debate and political conflict—in the context of chaotic political, financial, and social situations—rather than the outcome of academic arguments. It is worth noting, however, that a number of US economists of the time, e.g., Irving Fisher, then president of the American Economic Association, repeatedly argued that concentration of income and wealth was becoming as dangerously excessive in America as it had been for a long time in Europe, and called for steep tax progressivity (see e.g., Fisher, 1919). It is equally difficult to know whether economists had a major impact on the great reversal in top tax rates that occurred in the 1970–1980s during the Thatcher and Reagan conservative revolutions in Anglo-Saxon countries. The influential literature showing that top tax rate cuts can generate large responses of reported taxable income came after top tax rate cuts (e.g., Feldstein, 1995).

Today, most governments also draw on the work of commissions, panels, or reviews to justify tax and transfer reforms. Such reviews often play a big role in the public debate. They are sometimes commissioned by the government itself (e.g., the President's Advisory Panel on Federal Tax Reform in the United States, US Treasury, 2005), by independent policy research institutes (e.g., the Mirrlees review on Reforming the Tax System for the 21st Century in the United Kingdom, Mirrlees (2010,2011)), or proposed by independent academics (e.g., Landais et al., 2011 for France). Such reviews always involve tax scholars who draw on the academic economic literature to shape their recommendations.[19] The press also consults tax scholars to judge the merits of reforms proposed by politicians, and tax scholars naturally use findings from the academic literature when voicing their views.

2.2. History of the Field of Optimal Income Taxation

We offer here only a brief overview covering solely optimal income taxation.[20] The modern analysis of optimal income taxation started with Mirrlees (1971) who rigorously posed and solved the problem. He considered the maximization of a social welfare function based on individual utilities subject to a government budget constraint and

---

[17] See e.g., Mehrotra (2005) for a longer discussion of the role of Seligman on US tax policy at the beginning of the 20th century.

[18] This is particularly true in countries like France where mainstream laissez-faire economists had little sympathy for Anglo-Saxon utilitarian arguments, and were originally very hostile to tax progressivity, which they associated with radical utopia and with the French Revolution. See e.g., Delalande (2011a,b, pp. 166-170).

[19] Boadway (2012), Chapter 1 provides a longer discussion of the role played by such reviews.

[20] For a survey of historical fiscal doctrine in general see Musgrave (1985, chap. 1). For a more complete overview of modern optimal Boadway (2012), chapter 2.

incentive constraints arising from individuals' labor supply responses to the tax system.[21] Formally, in the Mirrlees model, people differ solely through their skill (i.e., their wage rate). The government wants to redistribute from high skill to low skill individuals but can only observe earnings (and not skills). Hence, taxes and transfers are based on earnings, leading to a non-degenerate equity-efficiency trade off.

Mirrlees (1971) had an enormous theoretical influence in the development of contract and information theory, but little influence in actual policy making as the general lessons for optimal tax policy were few. The most striking and discussed result was the famous zero marginal tax rate at the top. This zero-top result was established by Sadka (1976) and Seade (1977). In addition, if the minimum earnings level is positive with no bunching of individuals at the bottom, the marginal tax rate is also zero at the bottom (Seade, 1977). A third result obtained by Mirrlees (1971) and Seade (1982) was that the optimal marginal tax rate is never negative if the government values redistribution from high to low earners.

Stiglitz (1982) developed the discrete version of the Mirrlees (1971) model with just two skills. In this discrete case, the marginal tax rate on the top skill is zero making the zero-top result loom even larger than in the continuous model of Mirrlees (1971). That likely contributed to the saliency of the zero-top result. The discrete model is useful to understand the problem of optimal taxation as an information problem generating an incentive compatibility constraint for the government. Namely, the tax system must be set up so that the high skill type does not want to work less and mimic the low skill type. This discrete model is also widely used in contract theory and industrial organization. However, this discrete model has limited use for actual tax policy recommendations because it is much harder to obtain formulas expressed in terms of sufficient statistics or put realistic numbers in the discrete two skill model than in the continuous model.[22]

Atkinson and Stiglitz (1976) derived the very important and influential result that under separability and homogeneity assumptions on preferences, differentiated commodity taxation is not useful when earnings can be taxed nonlinearly. This famous result was influential both for shaping the field of optimal tax theory and in tax policy debates. Theoretically, it contributed greatly to shift the theoretical focus toward optimal nonlinear taxation and away from the earlier Diamond and Mirrlees (1971) model of differentiated commodity taxation (itself based on the original Ramsey (1927) contribution). Practically, it gave a strong rationale for eliminating preferential taxation of necessities on redistributive grounds, and using instead a uniform value-added-tax combined with income-based transfers and progressive income taxation. Even more importantly, the

---

[21] Vickrey (1945) had proposed an earlier formalization of the problem but without solving explicitly for optimal tax formulas.

[22] Stiglitz (1987, chap. 15) handbook chapter on optimal taxation provides a comprehensive optimal tax survey using the Stiglitz (1982) discrete model. In this chapter, we will not use the Stiglitz (1982) discrete model and present instead an alternative discrete model, first developed by Piketty (1997) which generates optimal tax formulas very close to those of the continuous model, and much easier to calibrate meaningfully.

Atkinson and Stiglitz (1976) result has been used to argue against the taxation of capital income and in favor of taxing solely earnings or consumption.

The optimal linear tax problem is technically simpler and it was known since at least Ramsey (1927) that the optimum tax rate can be expressed in terms of elasticities. Sheshinski (1972) is the first modern treatment of the optimal linear income tax problem. It was recognized early that labor supply elasticities play a key role in the optimal linear income tax rate. However, because of the disconnect between the nonlinear income tax analysis and the linear tax analysis, no systematic attempt was made to express nonlinear tax formulas in terms of estimable "sufficient statistics" until relatively recently.

Atkinson (1995), Diamond (1998), Piketty (1997), Saez (2001) showed that the optimal nonlinear tax formulas can also be expressed relatively simply in terms of elasticities.[23] This made it possible to connect optimal income tax theory to the large empirical literature estimating behavioral responses to taxation.

Diamond (1980) considered an optimal tax model with participation labor supply responses, the so-called extensive margin (instead of the intensive margin of the Mirrlees, 1971). He showed that the optimal marginal tax rate can actually be negative in that case. As we shall see, this model with extensive margins has received renewed attention in the last decade. Saez (2002a) developed simple elasticity-based formulas showing that a negative marginal tax rate (i.e., a subsidy for work) is optimal at the bottom in such an extensive labor supply model.

With hindsight, it may seem obvious that the quest for theoretical results in optimal income tax theory with broad applicability was doomed to yield only limited results. We know that the efficiency costs of taxation depend on the size of behavioral responses to taxes and hence that optimal tax systems are going to be heavily dependent on the size of those empirical parameters.

In this handbook chapter, in addition to emphasizing connections between theory and practical recommendations, we also want to flag clearly areas, where we feel that the theory fails to provide useful practical policy guidance. Those failures arise both because of limitations of empirical work and limitations of the theoretical framework. We discuss limitations of the standard utilitarian framework in Section 7. Another theoretical limitation arises because of behavioral considerations, i.e., the fact that individuals do not behave according to the standard utility maximization model, due to psychological effects and cognitive limitations. Such behavioral effects naturally affect the analysis and have generated an active literature both theoretical and empirical that we do not cover here (see e.g., Congdon, Mullainathan, & Schwartzstein, 2012 and the chapter by Chetty and Finkelstein in this volume for applications of behavioral economics to public economics).

---

[23] In the field of nonlinear pricing in industrial organization, the use of elasticity-based formulas came earlier (see e.g., Wilson, 1993).

3. CONCEPTUAL BACKGROUND

3.1. Utilitarian Social Welfare Objective

The dominant approach in normative public economics is to base social welfare on individual utilities. The simplest objective is to maximize the sum of individual utilities, the so-called utilitarian (or Benthamite) objective.[24]

Fixed Earnings. To illustrate the key ideas, consider a simple economy with a population normalized to one and an exogenous pre-tax earnings distribution with cumulative distribution function $H(z)$, i.e., $H(z)$ is the fraction of the population with pre-tax earnings below z. Let us assume that all individuals have the same utility function $u(c)$ increasing and concave in disposable income c (since there is only one period, disposable income is equal to consumption). Disposable income is pre-tax earnings minus taxes on earnings so that $c = z - T(z)$. The government chooses the tax function $T(z)$ to maximize the utilitarian social welfare function:

$$SWF = \int_0^\infty u(z - T(z))dH(z) \quad \text{subject to} \quad \int_0^\infty T(z)dH(z) \geq E \ (p),$$

where E is an exogenous revenue requirement for the government and p is the Lagrange multiplier of the government budget constraint. As incomes z are fixed, this is a point–wise maximization problem and the first order condition in $T(z)$ is simply:

$$u'(z - T(z)) = p \Rightarrow c = z - T(z) = \text{constant across } z.$$

Hence, utilitarianism with fixed earnings and concave utility implies full redistribution of incomes. The government confiscates 100% of earnings, funds its revenue requirement, and redistributes the remaining tax revenue equally across individuals. This result was first established by Edgeworth (1897). The intuition for this strong result is straightforward. With concave utilities, marginal utility $u'(c)$ is decreasing with c. Hence, if $c_1 < c_2$ then $u'(c_1) > u'(c_2)$ and it is desirable to transfer resources from the person consuming c_2 to the person consuming c_1.

Generalized social welfare functions of the form $\int G(u(c))dH(z)$ where $G(\cdot)$ is increasing and concave are also often considered. The limiting case where $G(\cdot)$ is infinitely concave is the *Rawlsian* (or *maxi-min*) criterion where the government's objective is to maximize the utility of the most disadvantaged person, i.e., maximize the minimum utility (maxi-min). In this simple context with fixed incomes, all those objectives also leads to 100% redistribution as in the standard utilitarian case.

Finally, with heterogeneous utility functions $u_i(c)$ across individuals, the utilitarian optimum is such that $u_i'(c)$ is constant over the population. Comparing the levels of marginal utility of consumption conditional on disposable income $z - T(z)$ across people with different preferences raises difficult issues of interpersonal utility comparisons.

---

[24] Utilitarianism as a social justice criterion was developed by the English philosopher Bentham in the late 18th century (Bentham, 1791).

There might be legitimate reasons, such as required health expenses due to medical conditions, that make marginal utility of consumption higher for some people than for others even conditional on after-tax income $z - T(z)$. Another legitimate reason would be the number of dependent children. Absent such need-based legitimate reasons, it does not seem feasible nor reasonable for society to discriminate in favor of those with high marginal utility of consumption (e.g., those who really enjoy consumption) against those with low marginal utility of consumption (e.g., those less able to enjoy consumption). This is not feasible because marginal utility of consumption cannot be observed and compared across individuals. Even if marginal utility were observable, it is unlikely that such discrimination would be acceptable to society (see our discussion in Section 6).

Therefore, it seems fair for the government to consider social welfare functions such that social marginal utility of consumption is the same across individuals conditional on disposable income. In the fixed earnings case, this means that the government can actually ignore individual utilities and use a "universal" social utility function $u(c)$ to evaluate social welfare. The concavity of $u(c)$ then reflects society's value for redistribution rather than directly individual marginal utility of consumption.[25] We will come back to this important point later on.

Endogenous Earnings. Naturally, the result of complete redistribution with concave utility depends strongly on the assumption of fixed earnings. In the real world, complete redistribution would certainly greatly diminish incentives to work and lead to a decrease in pre-tax earnings. Indeed, the goal of optimal income tax theory has been precisely to extend the basic model to the case with endogenous earnings (Vickrey, 1945 and Mirrlees, 1971). Taxation then generates efficiency costs as it reduces earnings, and the optimal tax problem becomes a non-trivial equity-efficiency trade off. Hence, with utilitarianism, behavioral responses are the sole factor preventing complete redistribution. In reality, society might also oppose complete redistribution on fairness grounds even setting aside the issue of behavioral responses. We come back to this limitation of utilitarianism in Section 6.

Let us therefore now assume that earnings are determined by labor supply and that individuals derive disutility from work. Individual i has utility $u^i(c, z)$ increasing in c but decreasing with earnings z. In that world, 100% taxation would lead everybody to completely stop working, and hence is not desirable.

Let us consider general social welfare functions of the type:

$$SWF = \int \omega_i G(u^i(c, z)) d\nu(i),$$

where $\omega_i \geq 0$ are *Pareto weights* independent of individual choices (c, z) and $G(\cdot)$ an increasing transformation of utilities, and $d\nu(i)$ is the distribution of individuals.

[25] Naturally, the two concepts are not independent. If individuals have very concave utilities, they will naturally support more redistribution under the "veil of ignorance," and the government choice for $u(c)$ will reflect those views.

The combination of arbitrary Pareto weights ω_i and a social welfare function $G(\cdot)$ allows us to be fully general for the moment. We denote by

$$g_i = \frac{\omega_i G'(u^i) u_c^i}{p}$$

the *social marginal welfare weight* on individual i, with p the multiplier of the government budget constraint.

Intuitively, g_i measures the dollar value (in terms of public funds) of increasing consumption of individual i by \$1. With fixed earnings, any discrepancy in the g_i's across individuals calls for redistribution as it increases social welfare to transfer resources from those with lower g_i's toward those with higher g_i's. Hence, absent efficiency concerns, the government should equalize all the g_i's.[26] With endogenous earnings, the g_i's will no longer be equalized at the optimum. As we shall see, social preferences for redistribution enter optimal tax formulas solely through the g_i weights.

Under the utilitarian objective, $g_i = u_c^i/p$ is directly proportional to the marginal utility of consumption. Under the Rawlsian criterion, all the g_i are zero, except for the most disadvantaged.

In the simpler case with no income effects on labor supply, i.e., where utility functions take the quasi-linear form $u^i(c, z) = v^i(c - h^i(z))$ with $v^i(\cdot)$ increasing and concave and $h^i(z)$ increasing and convex, the labor supply decision does not depend on non-labor income (see Section 3.3) and the average of g_i across all individuals is equal to one. This can be seen as follows. The government is indifferent between one more dollar of tax revenue and redistributing \$1 to everybody (as giving one extra dollar lump–sum does not generate any behavioral response). The value of giving \$1 extra to person i, in terms of public funds, is g_i so that the value of redistributing \$1 to everybody is $\int g_i dv(i)$.

3.2. Fallacy of the Second Welfare Theorem

The second welfare theorem seems to provide a strikingly simple theoretical solution to the equity-efficiency trade off. Under standard perfect market assumptions, the second welfare theorem states that any Pareto efficient outcome can be reached through a suitable set of lump-sum taxes that depend on exogenous characteristics of each individual (e.g., intrinsic abilities or other endowments or random shocks), and the subsequent free functioning of markets with no additional government interference. The logic is very simple. If some individuals have better earnings ability than others and the government wants to equalize disposable income, it is most efficient to impose a tax (or a transfer) based on earnings ability and then let people keep 100% of their actual earnings at the margin.[27]

---

[26] As we saw, under utilitarianism and concave and uniform utility functions across individuals, this implies complete equalization of post-tax incomes.

[27] In the model above, the government would impose taxes T_i based on the intrinsic characteristics of individual i but independent of the behavior of individual i so as to equalize all the g_i's across individuals (in the equilibrium where each individual chooses labor supply optimally given T_i).

In standard models, it is assumed that the government cannot observe earnings abilities but only realized earnings. Hence, the government has to base taxes and transfers on actual earnings only, which distort earnings and create efficiency costs. This generates an equity–efficiency trade off. This informational structure puts optimal tax analysis on sound theoretical grounds and connects it to mechanism design. While this is a theoretically appealing reason for the failure of the second welfare theorem, in our view, there must be a much deeper reason for governments to systematically use actual earnings rather than proxies for ability in real tax systems.

Indeed, standard welfare theory implies that taxes and transfers should depend on any characteristic correlated with earnings ability in the optimal tax system. If the characteristic is immutable, then average social marginal utilities across groups with different characteristics should be perfectly equalized. Even if the characteristic is manipulable, it should still be used in the optimal system (see Section 6.1). In reality, actual income tax or transfer systems depend on very few other characteristics than income. Those characteristics, essentially family situation or disability status, seem limited to factors clearly related to need.[28]

The traditional way to resolve this puzzle has been to argue that there are additional horizontal equity concerns that prevent the government from using non–income characteristics for tax purposes (see e.g., Atkinson and Stiglitz (1980) pp. 354–5). Recently, Mankiw and Weinzierl (2010) argue that this represents a major failure of the standard social welfare approach. This shows that informational concerns and observability is not the overwhelming reason for basing taxes and transfers almost exclusively on income. This has two important consequences.

First, finding the most general mechanism compatible with the informational set of the government—as advocated for example in the New Dynamic Public Finance literature (see Kocherlakota, 2010 for a survey)—might not be very useful for understanding actual tax problems. Such an approach can provide valuable theoretical insights and results but is likely to generate optimal tax systems that are so fundamentally different from actual tax systems that they are not implementable in practice. It seems more fruitful practically to assume instead exogenously that the government can only use a limited set of tax tools, precisely those that are used in practice, and consider the optimum within the set of real tax systems actually used. In most of this chapter, we therefore pursue this "simple tax structure" approach.[29]

Second, it would certainly be useful to make progress on understanding what concepts of justice or fairness could lead the government to use only a specific subset of taxes

---

[28] When incomes were not observable, archaic tax systems did rely on quasi-exogenous characteristics such as nobility titles, or land taxes based on rarely updated cadasters (Ardant, 1971). Ironically, when incomes become observable, such quasi-first best taxes were replaced by second-best income-based taxes.

[29] As mentioned above, the set of tools available changes over time. For example, individual incomes become observable only in modern economies.

and deliberately ignore other tools—such as taxes based on non–income characteristics correlated with ability—that would be useful to maximize standard utilitarian social welfare functions. We will come back to those important issues in Section 6.1 where we study tagging and in Section 7 where we consider alternatives to utilitarianism.

3.3. Labor Supply Concepts

In this chapter, we always consider a population of measure one of individuals. In most sections, individuals have heterogeneous preferences over consumption and earnings. Individual i utility is denoted by $u^i(c, z)$ and is increasing in consumption c and decreasing in earnings z as earnings require labor supply. Following Mirrlees (1971), in most models, heterogeneity in preferences is due solely to differences in wage rates w^i where utility functions take the form $u(c, z/w^i)$ where $l = z/w^i$ is labor supply needed to earn z. Our formulation $u^i(c, z)$ is more general and can capture both heterogeneity in ability as well as heterogeneity in preferences. As mentioned earlier, we believe that heterogeneity is an important element of the real world and optimal tax results should be reasonably robust to it.

To derive labor supply concepts, we consider a linear tax system with a tax rate τ combined with a lump sum demogrant R so that the budget constraint of each individual is $c = (1 - \tau)z + R$.

Intensive Margin. Let us focus first on the intensive labor supply margin, that is on the choice of how much to earn conditional on working. Individual i chooses z to maximize $u^i((1 - \tau)z + R, z)$ which leads to the first order condition

$$(1 - \tau)\frac{\partial u^i}{\partial c} + \frac{\partial u^i}{\partial z} = 0,$$

which defines implicitly the individual uncompensated (also called Marshallian) earnings supply function $z_u^i(1 - \tau, R)$.

The effect of $1 - \tau$ on z^i defines the *uncompensated elasticity* $e_u^i = \frac{1-\tau}{z_u^i}\frac{\partial z_u^i}{\partial(1-\tau)}$ of earnings with respect to the *net-of-tax rate* $1 - \tau$. The effect of R on z_u^i defines the *income effect* $\eta^i = (1 - \tau)\frac{\partial z_u^i}{\partial R}$. If leisure is a normal good, an assumption we make from now on, then $\eta^i \leq 0$ as receiving extra non–labor income induces the individual to consume both more goods and more leisure.

Finally, one can also define the compensated (also called Hicksian) earnings supply function $z_c^i(1 - \tau, u)$ as the earnings level that minimizes the cost necessary to reach utility u.[30] The effect of $1 - \tau$ on z^i keeping u constant defines the *compensated elasticity* $e_c^i = \frac{1-\tau}{z^i}\frac{\partial z_c^i}{\partial(1-\tau)}$ of earnings with respect to the *net-of-tax rate* $1 - \tau$. The compensated elasticity is always positive.

---

[30] Formally $z_c^i(1 - \tau, u)$ solves the problem $\min_z c - (1 - \tau)z$ subject to $u(c, z) \geq u$.

The *Slutsky equation* relates those parameters $e_c^i = e_u^i - \eta^i$. To summarize we have:

$$e_u^i = \frac{1-\tau}{z_u^i}\frac{\partial z_u^i}{\partial(1-\tau)} \gtrless 0, \quad \eta^i = (1-\tau)\frac{\partial z_u^i}{\partial R} \leq 0, \quad e_c^i = \frac{1-\tau}{z_c^i}\frac{\partial z_c^i}{\partial(1-\tau)} > 0,$$
$$\text{and} \quad e_c^i = e_u^i - \eta^i. \tag{1}$$

In the long-run process of development over the last century in the richest countries, wage rates have increased by a factor of five. Labor supply measured in hours of work has declined only very slightly (Ramey & Francis, 2009). If preferences for consumption and leisure have not changed, this implies that the uncompensated elasticity is close to zero. This does not mean however that taxes would have no effect on labor supply as a large fraction of taxes are rebated as transfers (see our discussion in Section 2). Therefore, on average, taxes are more similar to a compensated wage rate decrease than an uncompensated wage rate decrease. If income effects are large, government taxes and transfers could still have a large impact on labor supply.

Importantly, although we have defined those labor supply concepts for a linear tax system, they continue to apply in the case of a nonlinear tax system by considering the linearized budget at the utility maximizing point. In that case, we replace τ by the marginal tax rate $T'(z)$ and we replace R by virtual income defined as the non–labor income that the individual would get if her earnings were zero and she could stay on the virtual linearized budget. Formally $R = z - T(z) - (1 - T'(z)) \cdot z$.

Hence, the marginal tax rate $T'(z)$ reduces the marginal benefit of earning an extra dollar and reduces labor supply through *substitution* effects, conditional on the tax level $T(z)$. The income tax level $T(z)$ increases labor supply through *income* effects. In net, taxes (with $T'(z) > 0$ and $T(z) > 0$) hence have an ambiguous effect on labor supply while transfers (with $T'(z) > 0$ and $T(z) < 0$) have an unambiguously negative effect on labor supply.

Extensive Margin. In practice, there are fixed costs of work (e.g., searching for a job, finding alternative child care for parents, loss of home production, transportation costs, etc.). This can be captured in the basic model by assuming that choosing $z > 0$ (as opposed to $z = 0$) involves a discrete cost d_i.

It is possible to consider a pure extensive margin model by assuming that individual i can either not work (and earn zero) or work and earn z_i where z_i is fixed to individual i and reflects her earning potential. Assume that utility is linear, i.e., $u_i = c_i - d_i \cdot l_i$ where c_i is net-of-tax income, d_i is the cost of work and $l_i = 0, 1$ is a work dummy. In that case, individual i works if and only if $z_i - T(z_i) - d_i \geq -T(0)$, i.e., if $d_i \leq z_i - T(z_i) + T(0) = z_i \cdot (1 - \tau_p)$ where $\tau_p = [T(z_i) - T(0)]/z_i$. τ_p is the participation tax rate, defined as the fraction of earnings taxed when the individual goes from not working and earning zero to working and earning z_i. Therefore, the decision to work depends on the net-of-tax participation tax rate $1 - \tau_p$.

To summarize, there are three key concepts for any tax and transfer system $T(z)$. First, the transfer benefit with zero earnings $-T(0)$, sometimes called demogrant or lump-sum grant. Second, the marginal tax rate (or phasing-out rate) $T'(z)$: The individual keeps $1 - T'(z)$ for an additional \$1 of earnings. $1 - T'(z)$ is the key concept for the intensive labor supply choice. Third, the participation tax rate $\tau_p = [T(z) - T(0)]/z$: The individual keeps a fraction $1 - \tau_p$ of his earnings when going from zero earnings to earnings z. $1 - \tau_p$ is the key concept for the extensive labor supply choice. Finally, note that $T(z)$ integrates both the means-tested transfer program and the income tax that funds such transfers and other government spending. In practice transfer programs and taxes are often administered separately. The break even earnings point z^* is the point at which $T(z^*) = 0$. Above the break even point, $T(z) > 0$ which encourages labor supply through income effects. Below the break even point, $T(z) < 0$ which discourages labor supply through income effects.

Tax Reform Welfare Effects and Envelope Theorem. A key element of optimal tax analysis is the evaluation of the welfare effects of small tax reforms. Consider a nonlinear tax $T(z)$. Individual i chooses z to maximize $u^i(z - T(z), z)$, leading to the first order condition $u_c^i \cdot (1 - T'(z)) + u_z^i = 0$. Consider now a small reform $dT(z)$ of the nonlinear tax schedule. The effect on individual utility u^i is

$$du^i = u_c^i \cdot [-dT(z)] + u_c^i \cdot [1 - T'(z)]dz + u_z^i \cdot dz = u_c^i \cdot [-dT(z)],$$

where dz is the behavioral response of the individual to the tax reform and the second equality is obtained because of the first order condition $u_c^i \cdot (1 - T'(z)) + u_z^i = 0$. This is a standard application of the envelope theorem. As z maximizes utility, any small change dz has no first order effect on individual utility. As a result, behavioral responses can be ignored and the change in individual welfare is simply given by the *mechanical effect* of the tax reform on the individual budget multiplied by the marginal utility of consumption.

4. OPTIMAL LINEAR TAXATION

4.1. Basic Model

Linear labor income taxation simplifies considerably the exposition but captures the key equity-efficiency trade off. Sheshinski (1972) offered the first modern treatment of optimal linear income taxation following the nonlinear income tax analysis of Mirrlees (1971). Both the derivation and the optimal formulas are also closely related to the more complex nonlinear case. It is therefore pedagogically useful to start with the linear case where the government uses a linear tax at rate τ to fund a demogrant R (and additional non-transfer spending E taken as exogenous).[31]

[31] In terms of informational constraints, the government would be constrained to use linear taxation (instead of the more general nonlinear taxation) if it can only observe the amount of each earnings transaction but cannot observe the identity of individual earners. This could happen for example if the government can only observe the total payroll paid by each employer but cannot observe individual earnings perhaps because there is no identity number system for individuals.

Summing the Marshallian individual earnings functions $z_u^i(1-\tau, R)$, we obtain aggregate earnings which depend upon $1 - \tau$ and R and can be denoted by $Z_u(1 - \tau, R)$. The government's budget constraint is $R + E = \tau Z_u(1 - \tau, R)$, which defines implicitly R as a function of τ only (as we assume that E is fixed exogenously). Hence, we can express aggregate earnings as a sole function of $1 - \tau : Z(1 - \tau) = Z_u(1 - \tau, R(\tau))$. The tax revenue function $\tau \to \tau Z(1 - \tau)$ has an inverted U-shape. It is equal to zero both when $\tau = 0$ (no taxation) and when $\tau = 1$ (complete taxation) as 100% taxation entirely discourages labor supply. This curve is popularly called the Laffer curve although the concept of the revenue curve has been known since at least Dupuit (1844). Let us denote by $e = \frac{1-\tau}{Z} \frac{dZ}{d(1-\tau)}$ the elasticity of aggregate earnings with respect to the net-of-tax rate. The tax rate τ^* maximizing tax revenue is such that $Z(1 - \tau) - \tau \frac{dZ}{d(1-\tau)} = 0$, i.e., $\frac{\tau}{1-\tau} e = 1$. Hence, we can express τ^* as a sole function of e:

Revenue maximizing linear tax rate: $\quad \dfrac{\tau^*}{1 - \tau^*} = \dfrac{1}{e} \quad$ or $\quad \tau^* = \dfrac{1}{1 + e}$. $\qquad$ (2)

Let us now consider the maximization of a general social welfare function. The demogrant R evenly distributed to everybody is equal to $\tau Z(1 - \tau) - E$ and hence disposable income for individual i is $c^i = (1 - \tau)z^i + \tau Z(1 - \tau) - E$ (recall that population size is normalized to one). Therefore, the government chooses τ to maximize

$$SWF = \int_i \omega^i G[u^i((1 - \tau)z^i + \tau Z(1 - \tau) - E, z^i)]dv(i).$$

Using the envelope theorem from the choice of z^i in the utility maximization problem of individual i, the first order condition for the government is simply

$$0 = \frac{dSWF}{d\tau} = \int_i \omega^i G'(u^i)u_c^i \cdot \left[Z - z^i - \tau \frac{dZ}{d(1 - \tau)}\right]dv(i),$$

The first term in the square brackets $Z - z^i$ reflects the mechanical effect of increasing taxes (and the demogrant) absent any behavioral response. This effect is positive when individual income z^i is less than average income Z. The second term $-\tau \, dZ/d(1 - \tau)$ reflects the efficiency cost of increasing taxes due to the aggregate behavioral response. This is an efficiency cost because such behavioral responses have no first order positive welfare effect on individuals but have a first order negative effect on tax revenue.

Introducing the aggregate elasticity e and the "normalized" social marginal welfare weight $g^i = \omega^i G'(u^i)u_c^i / \int \omega^j G'(u^j)u_c^j dv(j)$, we can rewrite the first order condition as:

$$Z \cdot \left[1 - \frac{\tau}{1 - \tau}e\right] = \int_i g_i z_i dv(i).$$

Hence, we have the following optimal linear income tax formula

Optimal linear tax rate: $\quad \tau = \dfrac{1 - \bar{g}}{1 - \bar{g} + e} \quad$ with $\quad \bar{g} = \dfrac{\int g_i z_i dv(i)}{Z}$. $\qquad$ (3)

$\bar{g}$ is the average "normalized" social marginal welfare weight weighted by pre-tax incomes z_i. $\bar{g}$ is also the ratio of the average income weighted by individual social welfare weights g_i to the actual average income Z. Hence, $\bar{g}$ measures where social welfare weights are concentrated on average over the distribution of earnings. An alternative form for formula (3) often presented in the literature takes the form $\tau = -cov(g_i, z_i/Z)/[-cov(g_i, z_i/Z) + e]$ where $cov(g_i, z_i/Z)$ is the covariance between social marginal welfare weights g_i and normalized earnings z_i/Z. As long as the correlation between g_i and z_i is negative, i.e., those with higher incomes have lower social marginal welfare weights, the optimum τ is positive. Five points are worth noting about formula (3).

First, the optimal tax rate decreases with the aggregate elasticity e. This elasticity is a mix of substitution and income effects as an increase in the tax rate τ is associated with an increase in the demogrant $R = \tau Z(1-\tau) - E$. Formally, one can show that $e = [\bar{e}_u - \bar{\eta}]/[1 - \bar{\eta}\tau/(1-\tau)]$ where $\bar{e}_u = \frac{1-\tau}{Z_u} \frac{\partial Z_u}{\partial(1-\tau)}$ is the average of the individual uncompensated elasticities e_u^i weighted by income z^i and $\bar{\eta} = (1-\tau)\frac{\partial Z_u}{\partial R}$ is the unweighted average of individual income effects η^i.[32] This allows us to rewrite the optimal tax formula (3) in a slightly more structural form as $\tau = (1 - \bar{g})/(1 - \bar{g} - \bar{g} \cdot \bar{\eta} + \bar{e}_u)$.

When the tax rate maximizes tax revenue, we have $\tau = 1/(1 + e)$ and then $e = \bar{e}_u$ is a pure uncompensated elasticity (as the tax rate does not raise any extra revenue at the margin). When the tax rate is zero, e is conceptually close to a compensated elasticity as taxes raised are fully rebated with no efficiency loss.[33]

Second, the optimal tax rate naturally decreases with $\bar{g}$ which measures the redistributive tastes of the government. In the extreme case where the government does not value redistribution at all, $g_i \equiv 1$ and hence $\bar{g} = 1$ and $\tau = 0$ is optimal.[34] In the polar opposite case where the government is Rawlsian and maximizes the lump sum demogrant (assuming the worst-off individual has zero earnings), then $\bar{g} = 0$ and $\tau = 1/(1 + e)$, which is the revenue maximizing tax rate from Eq. (2). As mentioned above, in that case $e = \bar{e}_u$ is an uncompensated elasticity.

Third and related, for a given profile of social welfare weights (or for a given degree of concavity of the utility function in the homogeneous utilitarian case), the higher the pre-tax inequality at a given τ, the lower $\bar{g}$, and hence the higher the optimal tax rate. If there is no inequality, then $\bar{g} = 1$ and $\tau = 0$ with a lump-sum tax $-R = E$ is optimal. If inequality is maximal, i.e., nobody earns anything except for a single person who earns everything and has a social marginal welfare weight of zero, then $\tau = 1/(1 + e)$, again equal to the revenue maximizing tax rate.

Fourth, it is important to note that, as is usual in optimal tax theory, formula (3) is an implicit formula for τ as both e and especially $\bar{g}$ vary with τ. Under a standard utilitarian

---

[32] To see this, recall that $Z(1-\tau) = Z_u(1-\tau, \tau Z(1-\tau) - E)$ so that $\frac{dZ}{d(1-\tau)}\left[1 - \tau\frac{\partial Z_u}{\partial R}\right] = \frac{\partial Z_u}{\partial(1-\tau)} - Z\frac{\partial Z_u}{\partial R}$.

[33] It is not exactly a compensated elasticity as $\bar{e}_u$ is income weighted while $\bar{\eta}$ is not.

[34] This assumes that a lump sum tax E is feasible to fund government spending. If lump sum taxes are not feasible, for example because it is impossible to set taxes higher than earnings at the bottom, then the optimal tax in that case is the smallest τ such that $\tau Z(1-\tau) = E$, i.e., the level of tax required to fund government spending E.

social welfare criterion with concave utility of consumption, $\bar{g}$ increases with τ as the need for redistribution (i.e., the variation of the g_i with z_i) decreases with the level of taxation τ. This ensures that formula (3) generates a unique equilibrium for τ.

Fifth, formula (3) can also be used to assess tax reform. Starting from the current τ, the current estimated elasticity e, and the current welfare weight parameter $\bar{g}$, if $\tau < (1 - \bar{g})/(1 - \bar{g} + e)$ then increasing τ increases social welfare (and conversely). The tax reform approach has the advantage that it does not require knowing how e and $\bar{g}$ change with τ, since it only considers local variations.

Generality of the Formula. The optimal linear tax formula is very general as it applies to many alternative models for the income generating process. All that matters is the aggregate elasticity e and how the government sets normalized marginal welfare weights g^i. First, if the population is discrete, the same derivation and formula obviously apply. Second, if labor supply responses are (partly or fully) along the extensive margin, the same formula applies. Third, the same formula also applies in the long run when educational and human capital decisions are potentially affected by the tax rate as those responses are reflected in the long-run aggregate elasticity e (see e.g., Best & Kleven, 2012).[35]

Random Earnings. If earnings are generated by a partly random process involving luck in addition to ability and effort, as in Varian (1980) and Eaton and Rosen (1980), formula (3) still applies as long as the social welfare objective is defined over individual expected utilities.

To see this, suppose that pre-tax income for individual i is a random function of labor supply l^i and an idiosyncratic luck shock ε (with distribution dF^i) with $z^i = l^i + \varepsilon$ for simplicity. Individual i chooses l^i to maximize expected utility

$$EU^i = \int u^i((l^i + \varepsilon) \cdot (1 - \tau) + R, l^i) dF^i(\varepsilon),$$

so that l^i is function of $1 - \tau$ and R. The government budget implies again that $R = \tau Z - E$ so that Z is also a function of $1 - \tau$ as in the standard model (recall that $R = \tau Z(1 - \tau) - E$ is an implicit function of τ). The government then chooses τ to maximize $SWF = \int \omega^i G(EU^i) d\nu(i)$. This again leads to formula (3) with $\bar{g}$ the "normalized" average of $g^i = \omega^i G'(EU^i) u_c^i$ weighted by incomes z^i where now the average is taken as a double integral over both $dF^i(\varepsilon)$ and $d\nu(i)$.

Therefore, the random earnings model generates both the same equity-efficiency trade-off and the same type of optimal tax formula. This shows the robustness of the optimal linear tax approach. This robustness was not clearly apparent in the literature because of the focus on the nonlinear income tax case where the two models no longer deliver identical formulas.[36]

---

[35] Naturally, such long-run responses are challenging to estimate empirically as short-term comparisons around a tax reform cannot capture them.

[36] Varian (1980) analyzes the optimal nonlinear tax with random earnings.

Political Economy and Median Voter. The most popular model for policy decisions among economists is the median voter model. As is well known, the median voter theorem applies for unidimensional policies and where individual preferences are single-peaked with respect to this unidimensional policy. In our framework, the unidimensional policy is the tax rate τ (as the demogrant R is a function of τ). Each individual has single-peaked preferences about the tax rate τ as $\tau \to u^i((1 - \tau)z_i(1 - \tau) + \tau Z(1 - \tau), z_i(1 - \tau))$ is single-peaked with a peak such that $-z_i + Z - \tau \, dZ/d(1 - \tau) = 0$, i.e., $\tau_i = (1 - z_i/Z)/(1 - z_i/Z + e)$. Hence, the median voter is the voter with median income z_m. Recall that with single-peaked preferences, the median voter preferred tax rate is a Condorcet winner, i.e., wins in majority voting against any other alternative tax rate.[37] Therefore, the median voter equilibrium has:

$$\text{Median voter optimal tax rate:} \quad \tau_m = \frac{1 - z_m/Z}{1 - z_m/Z + e}. \tag{4}$$

The formula implies that when the median z_m is close to the average Z, the optimal tax rate is low because a linear tax rate achieves little redistribution (toward the median) and hence a lump-sum tax is more efficient.[38] In contrast, when the median z_m is small relative to the average, the tax rate τ_m gets close to the revenue maximizing tax rate $\tau^* = 1/(1 + e)$ from Eq. (2).

Formula (4) is a particular case of formula (3) where social welfare weights are concentrated at the median so that $\bar{g} = z_m/Z$. This shows that there is a tight connection between optimal tax theory and political economy. Political economy uses social welfare weights coming out of the political game process rather than derived from marginal utility of consumption as in the standard utilitarian tax theory but the structure of resulting tax formulas is the same (see Persson & Tabellini, 2002, chap. 24 for a comprehensive survey of political economy applied to public finance). We come back to the determination of social welfare weights in Section 6.

Finally and as caveats, note that the median voter theory applies only to unidimensional policies so that those results do not carry over to the nonlinear income tax case. The political economy literature has also shown that real world outcomes differ substantially from median voter predictions.

4.2. Accounting for Actual Tax Rates

As we saw in Section 2, tax to GDP ratios in OECD countries are between 30% and 45% and the more economically meaningful tax to national income ratios between 35% and 50%. Quantitatively, most estimates of aggregate elasticities of taxable income are

---

[37] To see this, if the alternative is $\tau' < \tau_m$, everybody below and including the median prefers τ_m to τ' so that τ_m wins. Conversely, if $\tau' > \tau_m$, everybody above and including the median prefers τ_m to τ' and τ_m still wins.

[38] Formula (4) shows that if $z_m > Z$, then a negative tax rate is actually optimal. Empirically however, it is always the case that $z_m < Z$.

between .1 and .4 with .25 perhaps being a reasonable estimate (see Saez, Slemrod, & Giertz, 2012 for a recent survey), although there remains considerable uncertainty about these magnitudes.[39]

Table 2 proposes simple illustrative calculations using the optimal linear tax rate formula (3). It reports combinations of τ and $\bar{g}$ in various situations corresponding to different elasticities e (across columns) and different social objectives (across rows). We consider three elasticity scenarios. The first one has $e = .25$ which is a realistic mid-range estimate (Saez et al., 2012, Chetty, 2012). The second has $e = .5$, a high range elasticity scenario. We add a third scenario with $e = 1$, an extreme case well above the current average empirical estimates.

Panel A considers the standard case where $\bar{g}$ is pinned down by a given social objective criterion and τ is then given by the optimal tax formula. The first row is the Rawlsian criterion (or revenue maximizing tax rate) with $\bar{g} = 0$. The second row is a utilitarian criterion with coefficient of relative risk aversion (CRRA) equal to one (social marginal welfare weights are proportional to $u_c = 1/c$ where $c = (1 - \tau)z + R$ is disposable income).[40] Chetty (2006) shows that a CRRA equal to one is consistent with empirical labor supply behavior and hence a reasonable benchmark. The third row is the median voter optimum with a median to average earnings ratio of 70% (corresponding approximately to the current US distribution based on individual adult earnings from the Current Population Survey in 2010). Panel B considers the inverse problem of determining the social preference parameter $\bar{g}$ for a given tax rate τ. The first row uses $\tau = 35\%$, corresponding to a low tax country such as the United States. The second row uses $\tau = 50\%$, corresponding to a high tax country such as a typical country from the European Union. Three points should be noted.

First, panel A shows that an empirically realistic elasticity $e = .25$ implies a revenue maximizing tax rate of 80% which is considerably higher than any actual average tax rate, even in the countries with the highest tax to GDP ratios, around 50%. The optimal tax rate under the utilitarian criterion with CRRA coefficient equal to one is 61%. The optimal tax rate for the median earner is $\tau = 55\%$ which corresponds to average tax rates in high tax countries. Correspondingly as shown in panel B, with $e = .25$, a tax rate of 35%, such as current US tax rates, would be optimal in a situation where $\bar{g} = 87\%$, i.e., with low redistributive tastes. A tax rate of 50% (as in a high tax country) would be optimal with $\bar{g} = 75\%$.

Second, a fairly high elasticity estimate of $e = .5$ would still generate a revenue maximizing tax rate of 67%, above current rates in any country. The median voter optimum tax

---

[39] Note however that the tax base tends to be smaller than national income as some forms of income (or consumption) are excluded from the tax base. Therefore, with existing tax bases, the tax rate needed to raise say 40% of national income, will typically be somewhat higher, perhaps around 50%.

[40] $\bar{g}$ is endogenously determined using the actual US earnings distribution and assuming that government required spending E (outside transfers) is 10% of total actual earnings. The distribution is for earnings of individuals aged 25 to 64 from the 2011 Current Population Survey for 2010 earnings.

Table 2 Optimal Linear Tax Rate Formula $\tau = (1 - g)/(1 - g + e)$

	Elasticity $e = .25$ (empirically realistic)		Elasticity $e = .5$ (high)		Elasticity $e = 1$ (extreme)	
	Parameter g (%) (1)	Tax rate τ (2)	Parameter g (%) (3)	Tax rate τ (4)	Parameter g (%) (5)	Tax rate τ (6)
A. Optimal linear tax rate τ						
Rawlsian revenue maximizing rate	0	80	0	67	0	50
Utilitarian (CRRA $= 1$, $u_c = 1/c$)	61	61	54	48	44	36
Median voter optimum ($z_{median}/z_{average} = 70\%$)	70	55	70	38	70	23
B. Revealed preferences g for redistribution						
Low tax country (US): Tax rate $\tau = 35\%$	87	35	73	35	46	35
High tax country (EU): Tax rate $\tau = 50\%$	75	50	50	50	0	50

Notes: This table illustrates the use of the optimal linear tax rate formula $\tau = (1 - g)/(1 - g + e)$ derived in the main text. It reports combinations of τ and g in various situations corresponding to different elasticities e (across columns) and different social objectives (across rows). Recall that g is the ratio of average earnings weighted by social marginal welfare weights to unweighted average earnings. Panel A considers the standard case where g is pinned down by a given social objective criterion and τ is then given by the optimal tax formula. The first row is the Rawlsian criterion (or revenue maximizing tax rate) with g = 0. The second row is a utilitarian criterion with coefficient of relative risk aversion (CRRA) equal to one (social marginal welfare weights are proportional to $u_c = 1/c$ where $c = (1 - \tau)z + R$ is disposable income). g is endogenously determined using the actual US earnings distribution and assuming that government required spending (outside transfers) is 10% of total earnings. The third row is the median voter optimum with a median to average earnings ratio of 70% (corresponding approximately to the current US situation). Panel B considers the inverse problem of determining the social preference parameter g for a given tax rate τ. The first row uses $\tau = 35\%$, corresponding to a low tax country such as the United States. The second row uses $\tau = 50\%$, corresponding to a high tax country such as a typical country from the European Union.

rate of 38% would actually be close to the current US tax rate in that situation. A high tax rate of 50% would be rationalized by $\bar{g} = .5$, i.e., fairly strong redistributive tastes. The utilitarian criterion also generates an optimal tax rate close to 50% in that elasticity scenario.

Third, in the unrealistically high elasticity scenario $e = 1$, the revenue maximizing rate is 50%, about the current tax rate in countries with the highest tax to GDP ratios. Hence, only in that case would social preferences for redistribution be approaching the polar Rawlsian case.

4.3. Tax Avoidance

As shown by many empirical studies (see Saez et al., 2012 for a recent survey), responses to tax rates can also take the form of tax avoidance. We can define tax avoidance as changes in reported income due to changes in the form of compensation but not in the total level of compensation. Tax avoidance opportunities typically arise when taxpayers can shift part of their taxable income into another form of income or another time period that receives a more favorable tax treatment.[41]

The key distinction between real and tax avoidance responses is that real responses reflect underlying, deep individual preferences for work and consumption while tax avoidance responses depend critically on the design of the tax system and the avoidance opportunities it offers. While the government cannot change underlying deep individual preferences and hence the size of the real elasticity, it can change the tax system to reduce avoidance opportunities.

A number of papers incorporate avoidance effects for optimal tax design. In this chapter, we adapt the simple modeling of Piketty, Saez, and Stancheva (2011) to the linear tax case so as to capture the key tradeoffs as simply and transparently as possible.[42]

We can extend the original model as follows to incorporate tax avoidance. Let us denote by y real income and by x sheltered income so that taxable income is $z = y - x$. Taxable income z is taxed at linear tax rate τ, while sheltered income x is taxed at a constant and linear tax rate t lower than τ. Individual i's utility takes the form:

$$u_i(c, y, x) = c - h_i(y) - d_i(x),$$

where $c = y - \tau z - tx + R = (1 - \tau)y + (\tau - t)x + R$ is disposable after tax-income. $h_i(y)$ is the utility cost of earning real income y, and $d_i(x)$ is the cost of sheltering an amount of income x. We assume a quasi-linear utility to simplify the derivations and eliminate cross-elasticity effects in real labor supply and sheltering decisions. We assume that both

---

[41] Examples of such avoidance/evasion are (a) reductions in current cash compensation for increased fringe benefits or deferred compensation such as stock-options or future pensions, (b) increased consumption within the firm such as better offices, vacation disguised as business travel, private use of corporate jets, etc., (c) re-characterization of ordinary income into tax favored capital income, and (d) outright tax evasion such as using off-shore accounts.

[42] Slemrod and Kopczuk (2002) endogenize avoidance opportunities in a multi-good model where the government selects the tax base. Finally, a large literature (surveyed in Slemrod and Yitzhaki (2002)) analyzes optimal policy design in the presence of tax evasion.

$h_i(\cdot)$ and $d_i(\cdot)$ are increasing and convex, and normalized so that $h_i'(0) = d_i'(0) = 0$. Individual utility maximization implies that

$$h_i'(y_i) = 1 - \tau \quad \text{and} \quad d_i'(x_i) = \tau - t,$$

so that y_i is an increasing function of $1 - \tau$ and x_i is an increasing function of the tax differential $\tau - t$. Aggregating over all individuals, we have $Y = Y(1 - \tau) = \int y_i(1 - \tau) \, dv(i)$ with real elasticity $e_Y = [(1 - \tau)/Y] dY/d(1 - \tau) > 0$ and $X = X(\tau - t) = \int x_i(\tau - t) dv(i)$ increasing in $\tau - t$. Note that $X(\tau - t = 0) = 0$ as there is sheltering only when $\tau > t$.

Hence aggregate taxable income $Z = Z(1 - \tau, t) = Y(1 - \tau) - X(\tau - t)$ is increasing in $1 - \tau$ and t. We denote by $e = [(1 - \tau)/Z] \partial Z/\partial(1 - \tau) > 0$ the total elasticity of taxable income Z with respect to $1 - \tau$ when keeping t constant. Note that $e = (Y/Z)e_Y + ((1 - \tau)/Z) dX/d(\tau - t) > (Y/Z)e_Y$. We immediately obtain the following optimal formulas.

Partial Optimum. For a given t, the tax rate τ maximizing tax revenue $\tau Z(1 - \tau, t) + tX(\tau - t)$ is

$$\tau = \frac{1 + t \cdot (e - (Y/Z)e_Y)}{1 + e}. \tag{5}$$

General Optimum. Absent any cost of enforcement, the optimal global tax policy (τ, t) maximizing tax revenue $\tau[Y(1 - \tau) - X(\tau - t)] + tX(\tau - t)$ is

$$t = \tau = \frac{1}{1 + e_Y}. \tag{6}$$

Four elements are worth noting about formulas (5) and (6).

First, if $t = 0$ then Eq. (5) becomes $\tau = 1/(1 + e)$ as in the standard model, Eq. (2). In the narrow framework where the tax system is taken as given (i.e., there is nothing the government can do about tax evasion and income shifting), and where sheltered income is totally untaxed, it is irrelevant whether the elasticity e arises from real responses or avoidance responses, a point made by Feldstein (1999).

Second however, if $t > 0$, then sheltering creates a "fiscal externality," as the shifted income generates tax revenue. In that case, Eq. (5) implies that τ is above the standard revenue maximization rate $1/(1 + e)$. As discussed earlier and as shown in the empirical literature (Saez et al., 2012), it is almost always the case that large short-term behavioral responses generated by tax changes are due to some form of income shifting or income retiming that generates fiscal externalities.

Third and most important, the government can improve efficiency and its ability to tax by closing tax avoidance opportunities (setting $t = \tau$ in our model), in which case the tax avoidance response becomes irrelevant and the real elasticity e_Y is the only factor

limiting tax revenue.[43] This strong result is obtained under the assumption that the tax avoidance opportunity arises solely from a poorly designed tax system that can be fixed at no cost.

Fourth and related, actual tax avoidance opportunities come in two varieties. Some are indeed pure creations of the tax system—such as the exemption of fringe benefits or tax exempt local government bonds—and hence could be entirely eliminated by reforming the tax system. In that case, t is a free parameter that the government can change at no cost as in our model. Yet other tax avoidance opportunities reflect real enforcement constraints that are costly—sometimes even impossible—for the government to eliminate. For example, it is very difficult for the government to tax income from informal businesses using only cash transactions, monitor perfectly consumption inside informal businesses, or fight offshore tax evasion.[44] The important policy question is then what fraction of the tax avoidance elasticity can be eliminated by tax redesign and tax enforcement effort.[45]

4.4. Income Shifting

The previous avoidance model assumed that shifting was entirely wasteful so that there was no reason for the government to set t lower than τ to start with. In reality, there are sometimes legitimate efficiency or distributional reasons why a government would want to tax different forms of income differently. On efficiency grounds, the classic Ramsey theory of optimal taxation indeed recommends lower tax rates on the most elastic goods or factors (Ramsey, 1927 and Diamond & Mirrlees, 1971).

Let us therefore extend our previous model by considering that there are two sources of income that we will call labor income and capital income for simplicity.[46] We follow again the simple modeling presented in Piketty et al. (2011). In this chapter, we focus solely on the static equilibrium and abstract from explicit dynamic considerations.[47] Labor income and capital income may respond to taxes differently and individuals can at some cost shift income from one form to the other. For example, small business owners can choose to pay themselves in the form of salary or business profits.

We assume that labor income z_L is taxed linearly at rate τ_L, while capital income z_K is taxed linearly at rate τ_K. True labor (respectively, capital) income is denoted by γ_L, (respectively, γ_K) while reported labor (respectively, capital) income is $z_L = \gamma_L - x$

---

[43] Kopczuk (2005) shows that the Tax Reform Act of 1986 in the United States, which broadened the tax base and closed loopholes did reduce the elasticity of reported income with respect to the net-of-tax rate.

[44] Offshore tax evasion is very difficult to fight from a single country's perspective but can be overcome with international coordination. This shows again that whether a tax avoidance/evasion opportunity can be eliminated depends on the institutional framework.

[45] Slemrod and Kopczuk (2002) present a model with costs of enforcement, where the government can adopt a broader tax base but where expanding the tax base is costly, to capture this trade-off theoretically.

[46] Other examples could be individual income vs. corporate income, or realized capital gains vs. ordinary income, or self-employment earnings vs. employee earnings.

[47] Chiappori (2008) propose an optimal tax analysis with shifting between capital and labor income in an OLG model.

(respectively, $z_K = y_K + x$) where x represents the amount of income shifting between the tax bases. Individual i has utility function:

$$u_i(c, y_L, y_K, x) = c - h_{Li}(y_L) - h_{Ki}(y_K) - d_i(x),$$

with $c = R + (1 - \tau_L)z_L + (1 - \tau_K)z_K = R + (1 - \tau_L)y_L + (1 - \tau_K)y_K + (\tau_L - \tau_K)x,$

where $h_{Li}(y_L)$ is the cost of producing labor income y_L, $h_{Ki}(y_K)$ is the cost of producing capital income y_K, and $d_i(x)$ is the cost of shifting income from the labor to the capital base. We assume that h_{Li}, h_{Ki}, and d_i are all convex. Note that $d_i(x) \geq 0$ is defined for both positive and negative x. We assume that $d_i(0) = 0$ and $d_i'(0) = 0$ and that $d_i'(x) \gtrless 0$ if and only if $x \gtrless 0$.[48] Individual utility maximization implies that

$$h_{Li}'(y_{Li}) = 1 - \tau_L, \quad h_{Ki}'(y_{Ki}) = 1 - \tau_K, \quad \text{and} \quad d_i'(x) = \tau_L - \tau_K,$$

so that y_{Li} is an increasing function of $1 - \tau_L$, y_{Ki} is an increasing function of $1 - \tau_K$, and x_i is an increasing function of the tax differential $\tau_L - \tau_K$. Aggregating over all individuals, we have $Y_L(1 - \tau_L) = \int y_{Li} d\nu(i)$ with real elasticity $e_L > 0$, $Y_K(1 - \tau_K) = \int y_{Ki} d\nu(i)$ with real elasticity $e_K > 0$, and $X(\tau_L - \tau_K) = \int x_i d\nu(i)$ increasing in $\Delta\tau = \tau_L - \tau_K$ with $X(0) = 0$. We can derive the revenue maximizing tax rates τ_L and τ_K in the following three cases:

No Income Shifting. If $X \equiv 0$, then $\tau_L = 1/(1 + e_L)$ and $\tau_K = 1/(1 + e_K)$.

Finite Shifting Elasticity. If $e_L < e_K$, we have: $1/(1 + e_L) \geq \tau_L > \tau_K \geq 1/(1 + e_K)$ (and conversely if $e_L > e_K$).

Infinite Shifting Elasticity. In the limit where X' is very large and real responses have finite elasticities e_L and e_K, then $\tau_L = \tau_K = 1/(1 + \bar{e})$ where $\bar{e} = (Y_L e_L + Y_K e_K)/(Y_L + Y_K)$ is the average real elasticity (weighted by income).

Those results have four notable implications. First, absent any shifting elasticity, there is no cross elasticity and we obtain the standard Ramsey inverse elasticity rule for each income factor.[49]

Second, the presence of shifting opportunities brings the optimal tax rates τ_L and τ_K closer together (relative to those arising under the inverse elasticity rule). When the shifting elasticity is large, optimal tax rates τ_L and τ_K should be close—even if the real elasticities e_L and e_K are quite different. Importantly, the presence of shifting does not necessarily reduce the ability of the government to tax but only alters the relative mix of tax rates. For example, in the case with infinite shifting, the optimum tax rates on labor and capital are equal and should be based on the average of the real elasticities.

---

[48] This model nests the pure tax avoidance model of the previous section in the case where $y_K \equiv 0$, i.e., there is no intrinsic capital income.

[49] As we have no income effects, the elasticities are also compensated elasticities.

Third, in this simple model, deciding whether labor or capital income should be taxed more requires comparing the elasticities e_L and e_K of *real* labor and capital income, and not the elasticities of reported labor and capital income. Empirically, this would require changing simultaneously and equally both τ_L and τ_K to determine which factor responds most keeping the level of income shifting $x(\Delta\tau)$ constant. Concretely, if shifting elasticities are large, a cut in τ_K will produce a large response of reported capital income but at the expense of labor income. It would be wrong to conclude that τ_K should be reduced. It should instead be brought closer to τ_L.

Fourth, it is possible to consider a standard social welfare maximization objective. In that case, optimal tax rates depend also on the distribution of each form of income. For example, under a standard utilitarian criterion with concave social marginal utility of consumption, if capital income is more concentrated than labor income, it should be taxed more (everything else equal). Those distributive effects in optimal tax formulas are well known from the theory of optimal commodity taxation (Diamond, 1975; Diamond & Mirrlees, 1971).[50]

5. OPTIMAL NONLINEAR TAXATION

Formally, the optimal nonlinear tax problem is easy to pose. It is the same as the linear tax problem except that the government can now choose any nonlinear tax schedule $T(z)$ instead of a single linear tax rate τ with a demogrant R. Therefore, the government chooses $T(z)$ to maximize

$$SWF = \int_i \omega^i G(u^i(z^i - T(z^i), z^i))d\nu(i) \quad \text{subject to} \quad \int_i T(z^i)d\nu(i) \geq E\,(p),$$

and the fact that z^i is chosen by individual i to maximize her utility $u^i(z^i - T(z^i), z^i)$. Note that transfers and taxes are fully integrated. Those with no earnings receive a transfer $-T(0)$. We start the analysis with the optimal top tax rate. Next, we derive the optimal marginal tax rate at any income level z. Finally, we focus on the bottom of the income distribution to discuss the optimal profile of transfers.

In this chapter, we purposefully focus on intuitive derivations using small reforms around the optimum. This allows us to understand the key economic mechanisms and obtain formulas directly expressed in terms of estimable "sufficient statistics" (Chetty, 2009a; Saez, 2001). Hence, we will omit discussions of technical issues about regularity conditions needed for the optimal tax formulas.[51]

---

[50] Note that there also exists dynamic reasons—e.g., the relative importance of inheritance and life-cycle saving in aggregate wealth accumulation—explaining why one might want to tax capital income more than labor income. See Piketty and Saez (2012a).

[51] The optimal income tax theory following Mirrlees (1971) has devoted substantial effort studying those issues thoroughly (see e.g., Mirrlees (1976,1986, chap. 24) for extensive surveys). The formal derivations are gathered in the appendix.
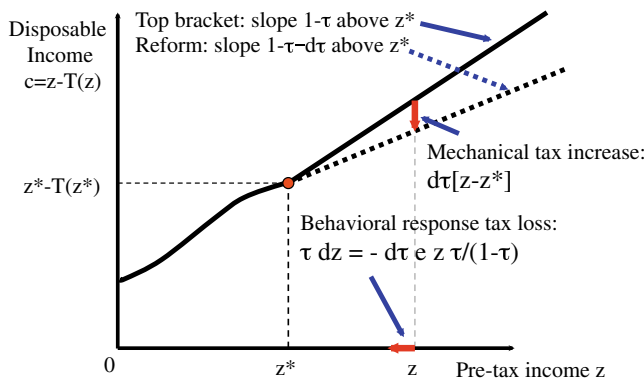
Figure 3 Optimal top tax rate derivation. The figure adapted from Diamond and Saez (2011), depicts the derivation of the optimal top tax rate $\tau = 1/(1 + ae)$ by considering a small reform around the optimum which increases the top marginal tax rate τ by $d\tau$ above z^*. A taxpayer with income z mechanically pays $d\tau[z - z^*]$ extra taxes but, by definition of the elasticity e of earnings with respect to the net-of-tax rate $1 - \tau$, also reduces his income by $dz = -d\tau ez/(1 - \tau)$ leading to a loss in tax revenue equal to $d\tau ez\tau/(1 - \tau)$. Summing across all top bracket taxpayers and denoting by z the average income above z^* and $a = z/(z - z^*)$, we obtain the revenue maximizing tax rate $\tau = 1/(1 + ae)$. This is the optimum tax rate when the government sets zero marginal welfare weights on top income earners.

5.1. Optimal Top Tax Rate

As discussed extensively in Section 2, the taxation of high income earners is a very important aspect of the tax policy debate. Initial progressive income tax systems were typically limited to the top of the distribution. Today, because of large increases in income concentration in a number of countries and particularly the United States (Piketty & Saez, 2003), the level of taxation of top incomes (e.g., the top 1%) matters not only for symbolic equity reasons but also for quantitatively for revenue raising needs.

5.1.1. Standard Model

Let us assume that the top tax rate above a fixed income level z^* is constant and equal to τ as illustrated on Figure 3. Let us assume that a fraction q of individuals are in the top bracket. To obtain the optimal τ, we consider a small variation $d\tau$ as depicted on Figure 3. Individual i earning z^i above z^*, mechanically pays $[z^i - z^*]d\tau$ extra in taxes. This extra tax payment creates a social welfare loss (expressed in terms of government public funds) equal to $-g^i \cdot [z^i - z^*]d\tau$ where $g^i = \omega_i G'(u^i)u_c^i/p$ is the social marginal welfare weight on individual i. [52] Finally, the tax change triggers a behavioral response dz^i leading to an additional change in taxes τdz^i. Using the elasticity of reported income z^i with respect to the net–of–tax rate $1 - \tau$, we have $dz^i = -e^i z^i d\tau/(1 - \tau)$. Hence, the

---

[52] Because the individual chooses z^i to maximize utility, the money-metric welfare effect of the reform on individual i is given by $[z^i - z^*]d\tau$ using the standard envelope theorem argument (see the end of Section 3.3).

net effect of the small reform on individual i is:

$$\left[(1 - g^i)(z^i - z^*) - e^i z^i \frac{\tau}{1 - \tau}\right] d\tau.$$

To obtain the total effect on social welfare, we simply aggregate the welfare effects across all top bracket taxpayers so that we have:

$$dSWF = \left[(1 - g)(z - z^*) - ez\frac{\tau}{1 - \tau}\right] q \, d\tau,$$

where q is the fraction of individuals in the top bracket, z is average income in the top bracket, g is the average social marginal welfare weight (weighted by income in the top bracket $z^i - z^*$) of top bracket individuals, and e is the average elasticity (weighted by income z^i) of top bracket individuals. We can introduce the tail–parameter $a = z/(z - z^*)$ to rewrite $dSWF$ as

$$dSWF = \left[1 - g - a \cdot e \frac{\tau}{1 - \tau}\right] (z - z^*) q \, d\tau.$$

At the optimum, $dSWF = 0$, leading to the following optimal top rate formula.

Optimal top tax rate: $\tau = \dfrac{1 - g}{1 - g + a \cdot e}.$ (7)

Formula (7) expresses the optimal top tax rate in terms of three parameters: a parameter g for social preferences, a parameter e for behavioral responses to taxes, and a parameter a for the shape of the income distribution.[53] Five points are worth noting about formula (7).

First, the optimal tax rate decreases with g, the social marginal welfare weight on top bracket earners. In the limit case where society does not put any value on the marginal consumption of top earners, the formula simplifies to $\tau = 1/(1 + a \cdot e)$ which is the revenue maximizing top tax rate. A utilitarian social welfare criterion with marginal utility of consumption declining to zero, the most commonly used specification in optimal tax models following Mirrlees (1971), has the implication that g converges to zero when z^* grows to infinity.

Second, the optimal tax rate decreases with the elasticity e as a higher elasticity leads to larger efficiency costs. Note that this elasticity is a mixture of substitution and income effects as an increase in the top tax rate generates both substitution and income effects.[54] Importantly, for a given compensated elasticity, the presence of income effects *increases* the optimal top tax rate as raising the tax rate reduces disposable income and hence increases labor supply.

---

[53] Note that the derivation and formula are virtually the same as for the optimal linear rate by simply multiplying e by the factor $a > 1$. Indeed, when $z^* = 0$, $a = z/(z - z^*) = 1$ and the problem boils down to the optimal linear tax problem.

[54] Saez (2001) provides a decomposition and shows that $e = \bar{e}^u + \bar{\eta} \cdot (a - 1)/a$ with $\bar{e}^u$ the average (income weighted) uncompensated elasticity and $\bar{\eta}$ the (unweighted) average income effect.
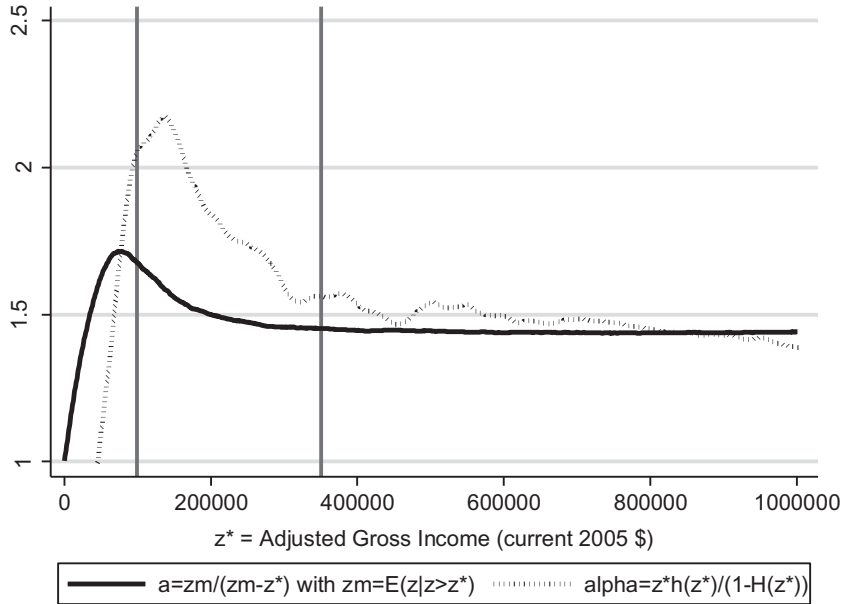
Figure 4 Empirical Pareto coefficients in the United States, 2005. The figure, from Diamond and Saez (2011), depicts in solid line the ratio $a = z_m/(z_m - z^*)$ with z^* ranging from \$0 to \$1,000,000 annual income and z_m the average income above z^* using US tax return micro data for 2005. Income is defined as Adjusted Gross Income reported on tax returns and is expressed in current 2005 dollars. Vertical lines depict the 90th percentile (\$99,200) and 99th percentile (\$350,500) nominal thresholds as of 2005. The ratio a is equal to one at $z^* = 0$, and is almost constant above the 99th percentile and slightly below 1.5, showing that the top of the distribution is extremely well approximated by a Pareto distribution for purposes of implementing the optimal top tax rate formula $\tau = 1/(1 + ae)$. Denoting by $h(z)$ the density and by $H(z)$ the cdf of the income distribution, the figure also displays in dotted line the ratio $\alpha(z^*) = z^* h(z^*)/(1 - H(z^*))$ which is also approximately constant, around 1.5, above the top percentile. A decreasing (or constant) $\alpha(z)$ combined with a decreasing $g^+(z)$ and a constant $e(z)$ implies that the optimal marginal tax rate $T'(z) = [1 - g^+(z)]/[1 - g^+(z) + \alpha(z)e(z)]$ increases with z.

Third, the optimal tax rate decreases with the parameter $a \geq 1$ which measures the thinness of the top tail of the income distribution. Empirically, $a = z/(z - z^*)$ is almost constant as z^* varies in the top tail of the earnings distribution. Figure 4 depicts a (as a function of z^*) for the case of the US pre-tax income distribution and shows that it is extremely stable above $z^* = \$400,000$, approximately the top 1% threshold.[55] This is due to the well-known fact—since at least Pareto (1896)—that the top tail is very closely approximated by a Pareto distribution.[56]

---

[55] This graph is taken from Diamond and Saez (2011) who use the 2005 distribution of total pre-tax family income (including capital income and realized capital gains) based on tax return data.

[56] A Pareto distribution with parameter a has a distribution of the form $H(z) = 1 - k/z^a$ and density $h(z) = ka/z^{1+a}$ (with k a constant parameter). For any z^*, the average income above z^* is equal to $z^* \cdot a/(a - 1)$.

Fourth and related, the formula shows the limited relevance of the zero–top tax rate result. Formally, z/z^* reaches 1 when z^* reaches the level of income of the single highest income earner, in which case $a = z/(z - z^*)$ is infinite and indeed $\tau = 0$, which is the famous zero–top rate result first demonstrated by Sadka (1976) and Seade (1977). However, notice that this result applies only to the very top income earner. Its lack of wider applicability can be verified empirically using distributional income tax statistics as we did in Figure 4 (see Saez, 2001 for an extensive analysis). Furthermore, under the reasonable assumption that the level of top earnings is not known in advance and where potential earnings are drawn randomly from an underlying Pareto distribution then, with the budget constraint satisfied in expectation, formula (7) remains the natural optimum tax rate (Diamond & Saez, 2011). This finding implies that the zero–top rate result and its corollary that marginal tax rates should decline at the top have no policy relevance.

Fifth, the optimal top tax rate formula is fairly general and applies equally to populations with heterogeneous preferences, discrete populations, or continuous populations. Although the optimal formula does not require the strong homogeneity assumptions of the Mirrlees (1971) problem, it is also the asymptotic limit of the optimal marginal tax rate of the fully nonlinear tax problem of Mirrlees (1971) as we shall see below.

5.1.2. Rent-Seeking Effects

Pay may not be equal to the marginal economic product for top income earners. In particular, executives can be overpaid if they are entrenched and can use their power to influence compensation committees. Indeed, a large literature in corporate finance has made those points (see for instance Bebchuk and Fried (2004) for an overview).[57]

There is relatively little work in optimal taxation that uses models where pay differs from marginal product.[58] Here we adapt the very basic model of Piketty et al. (2011) to illustrate the key issues created by rent seeking effects. Rothschild and Scheuer (2011) consider a more elaborate model with rent-seeking and earnings heterogeneity with two sectors where rent-seeking activities prone to congestion are limited to a single sector.[59]

Let us assume that individual i receives a fraction η of her actual product y. Individual i can exert productive effort to increase y or bargaining effort to increase η. Both types

---

[57] In principle, executives could also be underpaid relative to their marginal product if there is social outrage about high levels of compensation. In that case, a company might find it more profitable to under-pay its executives than face the wrath of its other employees, customers, or the public in general.

[58] A few studies have analyzed optimal taxation in models with labor market imperfections such as search models, union models, efficiency wages models (see Sorensen, 1999 for a survey). Few papers have addressed redistributive optimal tax policy in models with imperfect labor markets. Hungerbuhler, Lehmann, Parmentier Der Linden, and Bruno (2006) analyze a search model with heterogeneous productivity, and Stantcheva (2011) considers contracting models where firms cannot observe perfectly the productivity of their employees.

[59] In their model (and in contrast to the simple model we use here), when rent-seekers "steal" only from other rent-seekers, it is not optimal to impose high top tax rates because low top tax rates stimulate rent-seeking efforts, thereby congesting the rent-seeking sector and discouraging further entry.

of effort are costly to the individual. Hence, individual i utility is given by

$$u^i(c, \eta, y) = c - h_i(y) - k_i(\eta),$$

where c is disposable after-tax income, $h_i(y)$ is the cost of producing output y as in the standard model, and $k_i(\eta)$ is the cost of bargaining to get a share η of the product. Both h_i and k_i are increasing and convex.

Let $b = (\eta - 1)y$ be bargained earnings defined as the gap between received earnings ηy and actual product y. Note that the model allows both overpay (when $\eta > 1$ and hence $b > 0$) and underpay (when $\eta < 1$ and hence $b < 0$). Let us denote by $E(b)$ the average bargained earnings in the economy. In the aggregate, it must be the case that aggregate product must be equal to aggregate compensation. Hence, if $E(b) > 0$, average overpay $E(b)$ must come at the expense of somebody. Symmetrically, if $E(b) < 0$, average underpay $-E(b)$ must benefit somebody. For simplicity, we assume that any gain made through bargaining comes at the expense of everybody else in the economy uniformly. Hence, individual incomes are all reduced by the same amount $E(b)$ (or increased by $-E(b)$ if $E(b) < 0$).[60]

Because the government uses a nonlinear income tax schedule, it can adjust the demogrant intercept $-T(0)$ to fully offset $E(b)$. Effectively, the government can always tax (or subsidize) $E(b)$ at 100% before applying its nonlinear income tax. Hence, we can assume without loss of generality that the government absorbs one-for-one any change in $E(b)$. Therefore, we can simply define earnings as $z = \eta y = y + b$ and assume that those earnings are taxed nonlinearly.

Individual i chooses y and η to maximize:

$$u^i(c, \eta, y) = \eta \cdot y - T(\eta \cdot y) - h_i(y) - k_i(\eta),$$

which leads to the first order conditions

$$(1 - \tau)\eta = h_i'(y) \quad \text{and} \quad (1 - \tau)y = k_i'(\eta),$$

where $\tau = T'$ is the marginal tax rate. This naturally defines y_i, η_i as *increasing* functions of the net-of-tax rate $1 - \tau$. Hence $z_i = \eta_i \cdot y_i$ and $b_i = (\eta_i - 1) \cdot y_i$ are also functions of $1 - \tau$.

Let us consider as in the previous section the optimal top tax rate τ above income level z^*. We assume again that there is a fraction q of top bracket taxpayers. Let us denote by $z(1 - \tau)$, $y(1 - \tau)$, $b(1 - \tau)$ average reported income, productive earnings, and bargained earnings across all taxpayers in the top bracket. We can then define the real labor supply elasticity e_y and the total compensation elasticity e as:

$$e_y = \frac{1 - \tau}{y} \frac{dy}{d(1 - \tau)} \geq 0 \quad \text{and} \quad e = \frac{1 - \tau}{z} \frac{dz}{d(1 - \tau)} \geq 0.$$

---

[60] Piketty et al. (2011) show that this assumption can be relaxed without affecting the substance of the results.

We define s as the fraction of the marginal behavioral response due to bargaining and let $e_b = s \cdot e$ be the *bargaining elasticity component*:

$$s = \frac{db/d(1-\tau)}{dz/d(1-\tau)} = \frac{db/d(1-\tau)}{db/d(1-\tau) + d\gamma/d(1-\tau)} \quad \text{and} \quad e_b = s \cdot e = \frac{1-\tau}{z} \frac{db}{d(1-\tau)}.$$

This definition immediately implies that $(\gamma/z)e_b = (1-s) \cdot e$. By construction, $e = (\gamma/z)e_\gamma + e_b$. Importantly, s (and hence e_b) can be either positive or negative but it is always positive if individuals are overpaid (i.e., if $\eta > 1$). If individuals are underpaid (i.e., $\eta < 1$) then s (and hence e_b) may be negative.

For simplicity, let us assume that bargaining effects are limited to individuals in the top bracket. As there is a fraction q of top brackets individuals, we hence have $E(b) = qb(1-\tau)$. We assume that the government wants to maximize tax revenue collected from top bracket earners, taking into account bargaining effects:

$$T = \tau[\gamma(1-\tau) + b(1-\tau) - z^*]q - E(b) = \tau[\gamma(1-\tau) + b(1-\tau) - z^*]q - qb(1-\tau).$$

The second term $-E(b)$ arises because we assume that average underpay $-E(b)$ due to rent-seeking at the top is fully absorbed by the government budget as discussed above.

In this model, the top tax rate maximizing tax revenue satisfies the first order condition

$$0 = \frac{dT}{d\tau} = [\gamma + b - z^*]q - q\tau\frac{d\gamma}{d(1-\tau)} - q\tau\frac{db}{d(1-\tau)} + q\frac{db}{d(1-\tau)}.$$

The last term reflects the rent-seeking externality. Any decrease in top incomes due to a reduction in b creates a positive externality on all individuals, which can be recouped by the government by adjusting the demogrant. The optimal top tax rate can then be rewritten as follows:

Optimal top tax rate with rent-seeking: $\quad \tau^* = \dfrac{1 + a \cdot e_b}{1 + a \cdot e} = 1 - \dfrac{a(\gamma/z)e_\gamma}{1 + a \cdot e},$ (8)

τ^* decreases with the total e (keeping the bargaining component e_b constant) and increases with e_b (keeping e constant). It also decreases with the real elasticity e_γ (keeping e and γ/z constant) and increases with the level of overpayment $\eta = z/\gamma$ (keeping e_γ and e constant). If $e_\gamma = 0$ then $\tau^* = 1$. Two scenarios are theoretically possible.

Trickle–Up. In the case where top earners are overpaid relative to their productivity ($z > \gamma$), then $s > 0$ and hence $e_b > 0$ and the optimal top tax rate is higher than in the standard model (i.e., $\tau^* > 1/(1+a\cdot e)$). This corresponds to a "trickle–up" situation where a tax cut on upper incomes shifts economic resources away from the bottom and toward the top. Those effects can have a large quantitative impact on optimal top tax rates. In the extreme case where all behavioral responses at the top are due to rent-seeking effects ($e_b = e$ and $e_\gamma = 0$) then $\tau^* = 1$.

Trickle–Down. In the case where top earners are underpaid relative to their productivity ($z < \gamma$) it is possible to have $s < 0$ and hence $e_b < 0$, in which case the optimal top tax

rate is lower than in the standard model (i.e., $\tau^* < 1/(1 + a \cdot e)$). This corresponds to a "trickle-down" situation where a tax cut on upper incomes also shifts economic resources toward the bottom, as upper incomes are underpaid and hence work in part for the benefit of lower incomes.

Implementing formula (8) requires knowing not only how compensation responds to tax changes but also how real economic product responds to tax changes, which is considerably more difficult than estimating the standard taxable income elasticity e (see Piketty et al., 2011 for such an attempt). The issue of whether top earners deserve their incomes or are rent-seekers certainly looms large in the debate on top income taxation. Yet little empirical evidence can bear on the issue. This illustrates the limits of the theory of optimal taxation. Realistic departures from the standard economic model might be difficult to measure and yet can affect optimal tax rates in substantial ways.[61]

Finally, note that the model with rent-seeking is also related to the derivation of the optimal tax rates in the presence of externalities due to charitable giving responses (see e.g., Saez, 2004a) or the presence of transfers across agents (Chetty, 2009b).

5.1.3. International Migration

Taxes and transfers might affect migration in or out of the country. For example, high top tax rates might induce highly skilled workers to emigrate to low top tax rate countries.[62]

We consider a simplified version of the migration model of Mirrlees (1982) in order to obtain a simple formula.[63]

Let us assume that the only behavioral response to taxes is migration so that individual earnings z conditional on residence are fixed. Let us denote by $P(c|z)$ the number of resident individuals earning z when disposable domestic income is c. With the income tax, we have $c = z - T(z)$. We assume that $P(c|z)$ increases with c due to migration responses.

We can consider a small reform which increases taxes by dT for those earning z. The mechanical effect net of welfare is $dM + dW = (1 - g(z))P(c|z)dT$ where $g(z)$ is the social marginal welfare weight on individuals with earnings z. The net fiscal cost of somebody earning z emigrating is $T(z)$. We can define an elasticity of migration with respect to disposable income $\eta_m = [(z - T(z))/P(c|z)] \cdot \partial P/\partial c$. Hence the fiscal cost is $dB = -T(z) \cdot P(c|z) \cdot \eta_m/(z - T(z))$. Marginal emigrants are indifferent between emigrating or staying and hence the welfare cost is second order in this case as well. At the

---

[61] The same issue arises with optimal Ramsey taxation in the presence of imperfect competition, which has been explored in depth in the traditional optimal tax literature (see e.g., Auerbach and James (2002), Section 7 for a survey).

[62] The government can use other tools, such as immigration policy, to affect migration. Those other tools are taken here as given. Note that democracies typically do not control emigration but can control to some extent immigration. In the European Union context, emigration and immigration across EU countries is almost completely deregulated and hence our analysis is relevant in this context.

[63] Simula and Trannoy (2010) also derive optimal income tax formulas in a model including both migration and standard labor supply responses.

optimum, we have $dM + dW + dB = 0$, which implies:

$$\textbf{Optimal tax with migration only:} \quad \frac{T(z)}{z - T(z)} = \frac{1}{\eta_m} \cdot (1 - g(z)). \quad (9)$$

In the EU context, the most interesting application of the tax induced migration model is at the high income end. Indeed, there have been heated discussions of brain drain issues across EU countries due to differential tax rates at the top across countries. If we assume that high incomes respond both along the intensive margin as in Section 5.1.1 with elasticity e, and along the migration margin with elasticity η_m, then, it is possible to show that the optimal top rate maximizing tax revenue becomes (see Brewer, Shephard, & Saez, 2010):

$$\textbf{Optimal top tax rate adding migration effects:} \quad \tau^* = \frac{1}{1 + a \cdot e + \eta_m}. \quad (10)$$

For example if $a = 2$, $e = 0.25$, the optimal tax rate with no migration is $\tau^* = 1/(1 + 2 \cdot 0.25) = 2/3$. If there is migration with elasticity $\eta_m = 0.5$, then the optimal tax rate decreases to $\tau^* = 1/(1 + 2 \cdot 0.25 + 0.5) = 1/2$. Thus, large migration elasticities could indeed decrease significantly the ability of European countries to tax high incomes.

Two important additional points should be made. First, the size of the migration elasticity η_m depends not only on individual preferences but also on the size of the jurisdiction. Small jurisdictions—such as a town—typically have large elasticities as individuals can relocate outside the jurisdiction at low costs, for example without having to change jobs, etc. (see the chapter in this volume by Glaeser on urban public finance for a detailed discussion). The elasticity becomes infinite in the case of very small jurisdictions. Conversely, very large jurisdictions—such as a large country—have lower elasticities as it is costly to relocate. In the limit case of the full world, the migration elasticity is naturally zero. Therefore and as is well known, it is harder for small jurisdictions to implement redistributive taxation and indeed most redistributive tax and transfer programs tend to be carried out at the country level rather than the regional or city level.

Second and related, a single jurisdiction does not recognize the external cost it might impose on others by cutting its top tax rate. In that case, fiscal coordination across jurisdictions (e.g., European countries) could be mutually beneficial to internalize the externality. With complete fiscal coordination, the migration elasticity again becomes irrelevant for optimal tax policy (see the chapter by Keen and Konrad in this volume for an complete treatment of tax competition issues). When making policy recommendations, economists should try to be as clear as possible as to whether they are concerned with a single country optimum or with a global welfare perspective.[64]

---

[64] E.g. the Mirrlees Report is sometimes ambiguous as to whether the objective is to maximize social welfare at the global level or to find the tax system maximizing UK welfare.

5.1.4. Empirical Evidence on Top Incomes and Top Tax Rates

Micro-Level Tax Reform Studies. A very large literature has used tax reforms and micro-level tax return data to identify the elasticity of reported incomes with respect to the net-of-tax marginal rate. Those studies typically compare changes in pre-tax incomes of groups affected by a tax reform to changes in pre-tax incomes of groups unaffected by the reform. Hence, such tax reform-based analysis can only estimate short-term responses (typically 1–5 years) to tax changes. This literature, surveyed in Saez, Slemrod, and Giertz (2012), obtains three key conclusions that we briefly summarize here. First, there is substantial heterogeneity in the estimates: Many studies finding relatively small elasticity estimates (below 0.25), but some have found that tax reform episodes do generate large short-term behavioral responses, which imply large elasticities, particularly at the top of the income distribution. Second however, all the cases with large behavioral responses are due to tax avoidance such as retiming or income shifting. To our knowledge, none of the empirical tax reform studies to date have shown large responses due to changes in real economic behavior such as labor supply or business creation.[65] Furthermore, "anatomy analysis" shows that the large tax avoidance responses obtained are always the consequence of poorly designed tax systems offering arbitrage opportunities[66] or income retiming opportunities in anticipation of or just after-tax reforms.[67] When the tax system offers few tax avoidance opportunities, short-term responses to changes in tax rates are fairly modest with elasticities typically below 0.25.[68] Therefore, the results from this literature fit well with the tax avoidance model presented above with fairly small real elasticities and potentially large avoidance elasticities that can be sharply reduced through better tax design.

International Mobility. Mobility responses to taxation often loom larger in the policy debate on tax progressivity than traditional within country labor supply responses.[69] A large literature has shown that capital income mobility is a substantial concern (see e.g. the chapter by Keen and Konrad in this volume). However, there is much less empirical work on the effect of taxation on the spatial mobility of individuals, especially among high-skilled workers. A small literature has considered the mobility of people across local

---

[65] For example, the US Tax Reform Act of 1986 which cut the top marginal tax rate from 50% down to 28% led to a surge in reported top incomes but no effect on hours of work of top income earners (Moffitt & Wilhelm, 2000).

[66] For example, Slemrod (1996), Gordon and Slemrod (2000), and Saez (2004c) showed that part of the surge in top incomes immediately following the US tax cuts of the 1980s was due to income shifting from the corporate toward the individual sector.

[67] Auerbach (1988) showed that realized capital gains surged in 1986, in anticipation of the increase in the tax rate on realized capital gains starting in 1987. Goolsbee (2000) showed that stock-option realizations surged in 1992, in anticipation of the 1993 increase in top tax rates.

[68] For example, Kleven and Schultz (2012) provide very compelling estimates of modest—but not zero—elasticities around large tax reforms in Denmark, where the tax system offers few avoidance opportunities.

[69] For example, most of the objections in the popular and political debate to the recently proposed top marginal income tax rate of 75% in France are centered around mobility concerns: Will top talented workers (and top fortunes) leave France?

jurisdictions *within* countries.[70] While mobility costs within a country may be small, within country variations in taxes also tend to be modest. Therefore, it is difficult to extrapolate from those studies to international migration where both tax differentials and mobility costs are much higher. There is very little empirical work on the effect of taxation on international mobility partly due to lack of micro data with citizenship information and challenges in identifying causal tax effects on migration. In recent decades however, many countries, particularly in Europe, have introduced preferential tax rates for specific groups of foreign workers, and often highly paid foreign workers (see OECD, 2011c, chap. 4, Table 4.1, p. 138 for a summary of all such existing schemes). Such preferential tax schemes offer a promising route to identify tax induced mobility effects, recently exploited in two studies.

Kleven, Landais, and Saez (2013) study the tax induced mobility of professional football players in Europe and find substantial mobility elasticities. The mobility elasticity of the number of domestic players with respect to the domestic net-of-tax rate is relatively small, around .15. However, the mobility elasticity of the number of foreign players with respect to the net-of-tax rate that applies to foreign players is much larger, around 1. This difference is due to the fact that most players still play in their home country. Kleven et al. (in press) confirm that this latter result applies to the broader market of highly skilled foreign workers and not only football players. They show, in the case study of Denmark, that the preferential tax scheme for highly paid foreigners introduced in 1991 doubled the number of high earning foreigners in Denmark. This translates again into an elasticity of the number of foreign workers with respect to the net-of-tax rate above one.

Those results imply that, from a single country's perspective, as the number of foreigners at the top is still relatively small, the migration elasticity η_m of all top earners with respect to a single net-of-tax top rate is still relatively small, likely below .25 for most countries. This is the relevant elasticity to use in formula (10). Hence, the top income tax rate calculation is unlikely to be drastically affected by migration effects. However, this elasticity is likely to grow overtime as labor markets become better integrated and the fraction of foreign workers grows. Nevertheless, because the elasticity of the number of foreign workers with respect to the net-of-tax rate applying to foreign workers is so large, it is indeed advantageous from a single country perspective to offer such preferential tax schemes. This could explain why such schemes have proliferated in Europe in recent years. Such schemes are typical beggar-thy-neighbor policies which reduce the collective ability of countries to tax top earners. Hence, regulating such schemes at a supranational level (for example at the European Union level for European countries) is likely to become a key element in tax coordination policy debates.

---

[70] See Kirchgassner and Pommerehne (1996) on mobility across Swiss Cantons in response to Canton taxes or Young and Varner (2011) on mobility across US states in response to state income taxes.

Cross Country and Time Series Evidence. The simplest way to obtain evidence on the long-term behavioral responses of top incomes to tax rates is to use long time series analysis within a country or across countries. Data on top incomes overtime and across countries have been compiled by a number of recent studies (see Atkinson et al., 2011 for a survey) and gathered in the *World Top Incomes Database* (Alvaredo, Atkinson, Piketty & Saez 2011. A few recent studies have analyzed the link between top income shares and top tax rates (Atkinson & Leigh, 2010; Roine, Vlachos, & Waldenstrom, 2009; Piketty et al., 2011).

There is a strong negative correlation between top tax rates and top income shares, such as the fraction of total income going to the top 1% of the distribution. This long-run correlation is present overtime within countries as well as across countries. As an important caveat, the correlation between top tax rates and top income shares may not be causal as other policies potentially affecting top income shares, such as financial or industrial regulation or policies affecting Unions, may be correlated with top tax rate policy, creating an omitted variable bias. Alternatively and in reverse causality, higher top income shares may increase the political influence of top earners leading to lower top tax rates.[71]

Panel A in Figure 5 illustrates the cross-country evidence. It plots the change in top income shares from 1960–1964 to 2004–2009 (on the y-axis) against the change in the top marginal tax rate (on the x-axis) for 18 OECD countries. The figure shows a very clear and strong correlation between the cut in top tax rates and the increase in the top 1% income share with interesting heterogeneity. Countries such as France, Germany, Spain, Denmark, or Switzerland which did not experience any significant top rate tax cut did not experience large changes in top 1% income shares. Among the countries which experienced significant top rate cuts, some experience a large increase in top income shares (all five English speaking countries but also Norway and Finland) while others experience only modest increases in top income shares (Japan, Italy, Sweden, Portugal, and the Netherlands). Interestingly, no country experiences a significant increase in top income shares without implementing significant top rate tax cuts. Overall, the elasticity implied by this correlation is large, above 0.5. However, this evidence cannot tell whether the elasticity is due to real effects, tax evasion, or rent-seeking effects.

Panel B in Figure 5 illustrates the time series evidence for the case of the United States. It depicts the top 1% income shares including realized capital gains (pictured with full diamonds) and excluding realized capital gains (the empty diamonds) since 1913, which marks the introduction of the US federal income tax. Both top income shares, whether including or excluding realized capital gains, display an overall U-shape over the century. Panel A also displays (on the right y-axis) the federal individual income top marginal tax rate for ordinary income (dashed line), and for long-term realized capital gains (dotted line). Two important lessons emerge from this panel. Considering first the top income share excluding realized capital gains which corresponds roughly

---

[71] Analyzing the data in first differences can alleviate omitted variable bias but can only capture short-term effects of tax rates on top incomes, which might differ from long-term effects.
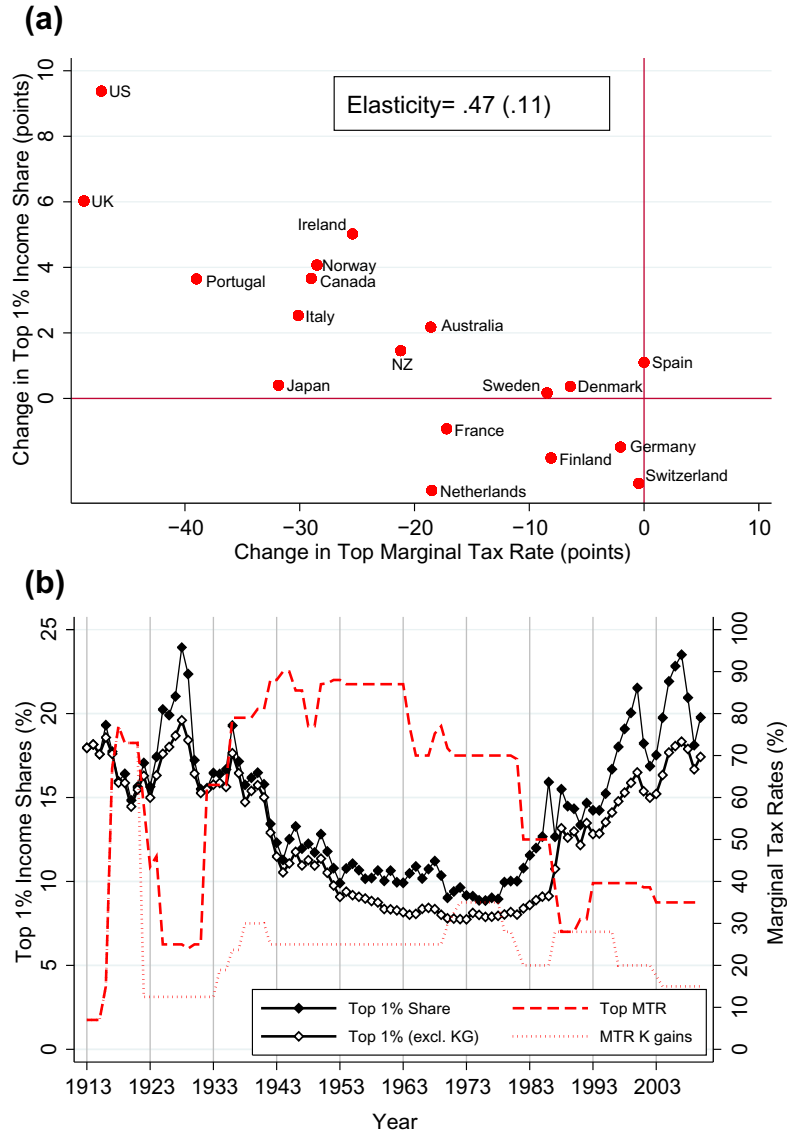
(a)

(b)

Figure 5 Top marginal tax rates and top incomes shares. This figure is from Piketty, Saez, and Stantcheva (2011). Panel A depicts the change in pre-tax top income shares against the change in pre-tax top income tax rate from 1960–1964 to 2005–2009 based on data for 18 OECD countries (exact years depend on availability of top income share data in the World Top Incomes Database (Alvaredo et al., 2011). Panel B depicts the pre-tax top 1% US income shares including realized capital gains in full diamonds and excluding realized capital gains in empty diamonds from 1913 to 2010. Computations are based on family market cash income. Income excludes government transfers and is before individual taxes (source is Piketty and Saez (2003), series updated to 2010). Panel B also depicts the top marginal tax rate on ordinary income and on realized long-term capital gains.

to income taxed according to the regular progressive schedule, there is a clear negative overall correlation between the top 1% income share and the top marginal tax rate, showing again that the elasticity of reported income with respect to the net-of-tax rate is large in the long run. Second, the correlation between the top 1% income share and the top tax rate also holds for the series including capital gains. Realized capital gains have been traditionally tax favored (as illustrated by the gap between the top tax rate and the tax rate on realized capital gains in the figure) and have constituted the main channel for tax avoidance of upper incomes.[72] This suggests that, in contrast to short-run tax reform analysis, income shifting responses cannot be the main channel creating the long-run correlation between top income shares and top tax rates.[73]

If the long-term correlation between top income shares and top tax rates is not driven by tax avoidance, the key question is whether it is driven by real supply side responses or whether it reflects rent-seeking effects whereby top earners can gain at the expense of others when top rates are low. In principle, the two types of behavioral responses can be distinguished by looking at economic growth as supply-side responses affect economic growth while rent-seeking responses do not. Piketty et al. (2011) analyze cross-country time series for OECD countries since 1960 and do not find any evidence that cuts in top tax rates stimulate growth. This suggests that rent-seeking effects likely play a role in the correlation between top tax rates and top incomes, and therefore that optimal top tax rates might be substantially larger than what it commonly assumed (say, above 80% rather than 50–60%). In our view, this is the right model to account for the quasi-confiscatory top tax rates during large parts of the 20th century (particularly in the US and in the UK; see Figure 1 above). Needless to say, more compelling empirical identification would be very useful to cast further light on this key issue for the optimal taxation of top earners.[74]

5.2. Optimal Nonlinear Schedule
5.2.1. Continuous Model of Mirrlees
It is possible to obtain the formula for the optimal marginal tax rate $T'(z)$ at income level z for the fully general nonlinear income tax using a similar variational method as the one used to derive the top income tax rate. To simplify the exposition, we consider the case with no income effects, where labor supply depends solely on the net-of-tax

---

[72] When individual top tax rates are high (relative to corporate and realized capital gains tax rates), it becomes more advantageous for upper incomes to organize their business activity using the corporate form and retain profits in the corporation. Profits only show up on individual returns as realized capital gains when the corporate stock is eventually sold (see Gordon and Slemrod, 2000 for a detailed empirical analysis).

[73] If top income share variations were due solely to tax avoidance, taxable income subject to the progressive tax schedule should be much more elastic than a broader income definition that also includes forms of income that are tax favored. Indeed, in the pure tax avoidance scenario, total real income of top earners should be completely inelastic to tax rates.

[74] Piketty et al. (2011) provide suggestive micro-level evidence. They show that CEO pay sensitivity to outcomes outside CEOs' control (such as industry wide shocks) is higher when top rates are low, both in the US time series and across countries.
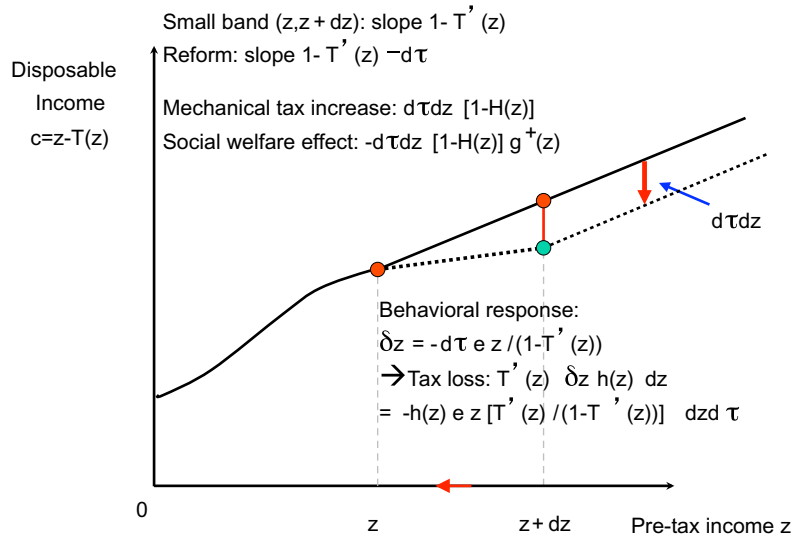
Figure 6 Derivation of the optimal marginal tax rate at income level *z*. The figure, adapted from Diamond and Saez (2011), depicts the optimal marginal tax rate derivation at income level *z* by considering a small reform around the optimum, whereby the marginal tax rate in the small band $(z, z + dz)$ is increased by $d\tau$. This reform mechanically increases taxes by $d\tau dz$ for all taxpayers above the small band, leading to a mechanical tax increase $d\tau dz[1 - H(z)]$ and a social welfare cost of $-d\tau dz[1 - H(z)]g^+(z)$. Assuming away income effects, the only behavioral response is a substitution effect in the small band: The $h(z)dz$ taxpayers in the band reduce their income by $\delta z = -d\tau ez/(1 - T'(z))$ leading to a tax loss equal to $-d\tau dzh(z)ezT'(z)/(1 - T'(z))$. At the optimum, the three effects cancel out leading to the optimal tax formula $T'(z)/(1 - T'(z)) = (1/e)(1 - g^+(z))(1 - H(z))/(zh(z))$, or equivalently $T'(z) = [1 - g^+(z)]/[1 - g^+(z) + \alpha(z)e]$ after introducing $\alpha(z) = zh(z)/(1 - H(z))$.

rate $1 - T'(z)$.[75] We present in the text a graphical proof adapted from Saez (2001) and Diamond and Saez (2011) and we relegate to the appendix the formal presentation and derivation in the standard Mirrlees model with no income effects (as in the analysis of Diamond, 1998).

Figure 6 depicts the optimal marginal tax rate derivation at income level z. Again, the horizontal axis in Figure 6 shows pre-tax income, while the vertical axis shows disposable income. Consider a situation in which the marginal tax rate is increased by $d\tau$ in the small band from z to $z + dz$, but left unchanged anywhere else. The tax reform has three effects.

First, the mechanical tax increase, leaving aside behavioral responses, will be the gap between the solid and dashed lines, shown by the vertical arrow equal to $dz\, d\tau$. The total mechanical tax increase is $dM = dz\, d\tau[1 - H(z)]$ as there are $1 - H(z)$ individuals above z.

---

[75] Atkinson (1995) and Diamond (1998) showed that this case generates simpler formulas. Saez (2001) considers the case with income effects.

Second, this tax increase creates a social welfare cost of $dW = -dz\, d\tau[1-H(z)]g^+(z)$ where $g^+(z)$ is defined as the average (unweighted) social marginal welfare weight for individuals with income above z.

Third, there is a behavioral response to the tax change. Those in the income range from z to $z + dz$ have a behavioral response to the higher marginal tax rate, shown by the horizontal line pointing left. Assuming away income effects, this is the only behavioral response; those with income levels above $z + dz$ face no change in marginal tax rates and hence have no behavioral response. A taxpayer in the small band reduces her income by $\delta z = -ez\, d\tau/(1 - T'(z))$ where e is the elasticity of earnings z with respect to the net-of-tax rate $1 - T'$. As there are $h(z)dz$ taxpayers in the band, those behavioral responses lead to a tax loss equal to $dB = -dz\, d\tau\, h(z)ezT'(z)/(1 - T'(z))$.[76]

At the optimum, the three effects should cancel out so that $dM + dW + dB = 0$. Define the local Pareto parameter as $\alpha(z) = zh(z)/(1 - H(z))$.[77] This leads to the following optimal tax formula

$$\textbf{Optimal nonlinear marginal tax rate:}\quad T'(z) = \frac{1 - g^+(z)}{1 - g^+(z) + \alpha(z) \cdot e} \qquad (11)$$

Formula (11) has essentially the same form as (7). Five further points are worth noting.

First, the simple graphical proof shows that the formula does not depend on the strong homogeneity assumptions of the standard Mirrlees model where individuals differ solely through a skill parameter. This implies that the formula actually carries over to heterogeneous populations as is the case of the basic linear tax rate formula (3).[78]

Second, the optimal tax rate naturally decreases with $g^+(z)$, the average social marginal welfare weight above z. Under standard assumptions where social marginal welfare weights decrease with income, $g^+(z)$ is decreasing in z. With no income effects, the average social marginal welfare weight is equal to one (see Section 3.1) so that $g^+(0) = 1$ and $g^+(z) < 1$ for $z > 0$. This immediately implies that $T'(z) \geq 0$ for any z, one of the few general results coming out of the Mirrlees model and first demonstrated by Mirrlees (1971) and Seade (1982).[79] A decreasing $g^+(z)$ tends to make the tax system

---

[76] This derivation has ignored the fact that the tax schedule is locally nonlinear. Saez (2001) shows that, in the exact formula for dB, the density $h(z)$ should be replaced by the "virtual density" $h^*(z)$ defined as the density at z that would arise if the nonlinear tax system were replaced by the linearized tax system at point z (see the appendix for a formal treatment).

[77] We call $\alpha(z)$ a local Pareto parameter because for an exact Pareto distribution, $\alpha(z)$ is constant and equal to the Pareto parameter a.

[78] This point does not seem to have been formally established in the case of optimal tax theory but is well known in the mathematically equivalent optimal nonlinear pricing problem in the Industrial Organization literature (see e.g., Wilson, 1993, Section 8.4).

[79] $T'(z) < 0$ is never optimal in the Mirrlees model when marginal welfare weights decrease with z. This is because increasing $T'(z)$ locally (as depicted on Figure 6) would raise more revenue from everybody above z which is desirable for redistribution. The behavioral response δz in the small band would further increase tax revenue (as $T'(z) < 0$) making the reform desirable.

more progressive. Note that the extreme Rawlsian case has $g^+(z) = 0$ for all z except at $z = 0$ (assuming realistically that the most disadvantaged are those with no earnings). In that case, the formula simplifies to $T'(z) = 1/(1 + \alpha(z) \cdot e)$ and the optimal tax system maximizes tax revenue raised to make the lump sum demogrant $-T(0)$ as large as possible.

Third, the optimal tax rate decreases with the elasticity e at income level z as a higher elasticity leads to larger efficiency costs in the small band $(z, z + dz)$. Note that this elasticity remains a pure substitution elasticity even in the presence of income effects.[80]

Fourth, the optimal tax rate decreases with the local Pareto parameter $\alpha(z) = zh(z)/[1 - H(z)]$ which reflects the ratio of the total income of those affected by the marginal tax rate at z relative to the number of people at higher income levels. The intuition for this follows the derivation from Figure 6. Increasing $T'(z)$ creates efficiency costs proportional to the number of people at income level z times the income level z while it raises more taxes (with no distortion) from everybody above z. As shown on Figure 4 for the US case, empirically $\alpha(z)$ first increases and then decreases before being approximately constant in the top tail. Hence, when z is large, formula (11) converges to the optimal top rate formula (7) that we derived earlier.

Fifth, suppose the government has no taste for redistribution and wants to raise an exogenous amount of revenue while minimizing efficiency costs. If lump sum taxes are realistically ruled out because those with no earnings could not possibly pay them, then the optimal tax system is still given by (11) with constant social marginal welfare weights and hence constant $g^+(z)$ set to exactly raise the needed amount of exogenous revenue (Saez, 1999, chap. 3).

Increasing Marginal Tax Rates at the Top. With an elasticity e constant across income groups, as $g^+(z)$ decreases with z and $\alpha(z)$ also decreases with z in the upper part of the distribution (approximately the top 5% in the US case, see Figure 4), formula (11 implies that the optimal marginal tax rate should increase with z at the upper end, i.e., the income tax should be progressive at the top. Diamond (1998) provides formal theoretical results in the Mirrlees model with no income effects.

Numerical Simulations. For low z, $g^+(z)$ decreases but $\alpha(z)$ increases. Numerical simulations calibrated using the actual US earnings distribution presented in Saez (2001) show that the $\alpha(z)$ effect dominates at the bottom so that the marginal tax rate is high and decreasing for low z. We come back to this important issue when we discuss the optimal profile of transfers below. Therefore, assuming that the elasticity is constant with z, the optimal marginal tax rate in the Mirrlees model is U-shaped with income, first decreasing with income and then increasing with income before converging to its limit value given by formula (7).

---

[80] Income effects positively affect labor supply above z so that the mechanical tax revenue increase is actually higher than $dz \, d\tau[1 - H(z)]$ and the optimal tax rate is correspondingly higher (see Saez, 2001).

5.2.2. Discrete Models

Stiglitz (1982) developed the 2 skill-type discrete version of the Mirrlees (1971) model where individuals can have either a low or a high wage rate. This discrete model has been used widely in the subsequent literature because it has long been perceived as more tractable than the continuous model of Mirrlees. However, the discrete model is perhaps deceiving when it comes to understanding optimal tax progressivity. Indeed, the zero top marginal tax rate result implies that the marginal tax rate on the highest skill is zero and hence lower than the marginal tax rate on the lowest skill, suggesting that the marginal tax rate should decrease with earnings. Furthermore, it is impossible to express optimal tax formulas in the Stiglitz (1982) model in terms of estimable statistics and hence to quantitatively calibrate the model.

More recently, Piketty (1997) introduced and Saez (2002a) further developed an alternative form of discrete Mirrlees model with a finite number of possible earnings levels $z_0 = 0 < z_1 < ... < z_N$ (corresponding for example to different possible jobs) but a continuum of individual types so that the fraction of individuals at each earnings level is a smooth function of the tax system. This model generates formulas close to the continuum case, and can also be easily extended to incorporate extensive labor supply responses, as we shall see.

Formally, individual i has a utility function $u^i(c_n, n)$ defined on after-tax income $c_n \geq 0$ and job choice $n = 0, ..., N$. Each individual chooses n to maximize $u^i(c_n, n)$ where $c_n = z_n - T_n$ is the after-tax reward in occupation n. For a given tax and transfer schedule $(c_0, ..., c_N)$, a fraction $h_n(c_0, ..., c_N)$ of individuals choose occupation n. It is assumed that the tastes for work embodied in the individual utilities are smoothly distributed so that the aggregate functions h_n are differentiable. Denoting by $n(i)$ the occupational choice of individual i, the government chooses $(T_0, ..., T_N)$ so as to maximize welfare

$$SWF = \int_i \omega_i G[u^i(z_{n(i)} - T_{n(i)}, n(i))]d\nu(i) \quad \text{s.t.} \quad \sum_n h_n T_n \geq E\ (p).$$

Even though the population is potentially very heterogeneous, as possible work outcomes are in finite number, the maximization problem is a simple finite dimensional maximization problem. The first order condition with respect to T_n is

$$(1 - g_n)h_n = \sum_{m=0}^{N} T_m \frac{\partial h_m}{\partial c_n} \quad \text{with} \quad g_n = \frac{1}{p\,h_n} \int_{i \in \text{job}\ n} \omega^i G'(u^i)u_c^i(c_n, n)d\nu(i). \quad (12)$$

Hence, g_n is the average social marginal welfare weight among individuals in occupation n.[81]

---

[81] When obtaining (12), it is important to note that, because of the envelope theorem, the effect of an infinitesimal change in c_n has no discrete effect on welfare for individuals moving in or out of occupation n. Hence, the welfare effects on movers is second order. See Saez (2002a), appendix for complete details.

This model allows for any type of behavioral responses. Two special cases are of particular interest: pure intensive responses as in the standard Mirrlees (1971) model and pure extensive responses. We consider in this section the intensive model case and defer to Section 5.3.2 the extensive model case.

The intensive model. The intensive model with no income effects (first developed by Piketty, 1997) can be obtained by assuming that the population is partitioned into N groups. An individual in group $n \in (0, \ldots, N-1)$ can only work in two adjacent occupations n and $n+1$. For example, with no effort the individual can hold job n and with some effort the individual can obtain job $n+1$.[82] This implies that the function h_n depends only on c_{n+1}, c_n, and c_{n-1}. Assuming no income effects, with a slight abuse of notation, h_n can be expressed as $h_n(c_{n+1}-c_n, c_n-c_{n-1})$. In that context, we can denote by $\tau_n = (T_n - T_{n-1})/(z_n - z_{n-1})$ the marginal tax rate between earnings levels z_{n-1} and z_n and by $e_n = \frac{1-\tau_n}{h_n} \frac{\partial h_n}{\partial(1-\tau_n)}$ the elasticity of the fraction of individuals in occupation n with respect to the net-of-tax rate $1 - \tau_n$. The optimal tax formula (12) can be rearranged as:

Optimal marginal tax rate, discrete model: $\quad \dfrac{\tau_n}{1-\tau_n} = \dfrac{1}{e_n}\left[\dfrac{\sum_{m \geq n}(1-g_m)h_m}{h_n} \right].$

$$(13)$$

The proof is presented in Saez (2002a). Note that the form of the optimal formula is actually very close the continuum case where the marginal tax rate from Eq. (11) can also be written as: $T'(z)/[1 - T'(z)] = (1/e)[\int_z^\infty (1 - g(z'))dH(z')/(zh(z))]$.

5.3. Optimal Profile of Transfers

5.3.1. Intensive Margin Responses

It is possible to obtain a formula for the optimal phase-out rate of the demogrant in the optimal income tax model of Mirrlees (1971) where labor supply responds only through the intensive margin.

Recall first that when the minimum income z_0 is positive, the optimal marginal tax rate at the very bottom is zero (this result was first proved by Seade, 1977). This can be seen from formula (11) as $G(z_0) = 1$.[83]

However, the empirically relevant case is $z_0 = 0$ with a non-zero fraction $h_0 > 0$ of the population not working and earning zero. In that case, the optimal phase-out rate τ_1

---

[82] Those preferences are embodied in the individual utility functions u^i. In the case just described, we would have $u^i(c, n) = c$, $u^i(c, n+1) = c - \theta_i$ with θ_i cost of effort to get job $n+1$, and $u^i(c, m) = -\infty$ if $m \notin \{n, n+1\}$.

[83] This result can be seen as the symmetric counterpart of the zero-top result. At the top, it is straightforward to show that the optimum marginal tax rate cannot be positive (if it were, set it to zero above z_{top}, the top earner works more, is better off, and pays the same taxes). However, it is not as easy to show that the top rate cannot be negative (this requires the more refined argument presented in comments of formula (11)). At the bottom symmetrically, it is straightforward to show that the optimum marginal tax rate cannot be negative (if it were, set it to zero below z_{bottom}, the bottom earner works less, is better off, and pays the same taxes). However, it is not as easy to show that the bottom rate cannot be positive (this again requires a symmetric argument to the one presented in comments of formula (11).)
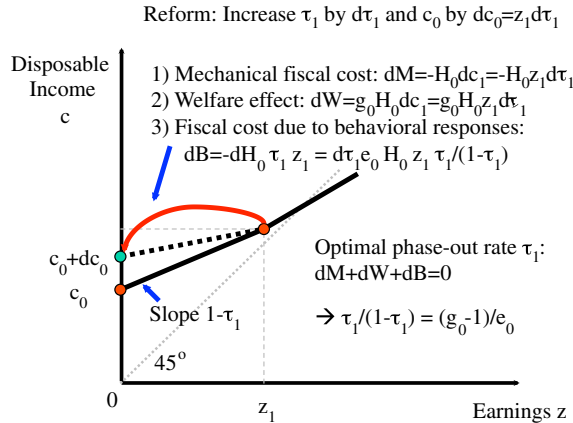
Reform: Increase τ_1 by $d\tau_1$ and c_0 by $dc_0=z_1 d\tau_1$

Disposable
Income
c

1) Mechanical fiscal cost: $dM=-H_0 dc_1=-H_0 z_1 d\tau_1$
2) Welfare effect: $dW=g_0 H_0 dc_1=g_0 H_0 z_1 d\tau_1$
3) Fiscal cost due to behavioral responses:
$dB=-dH_0 \tau_1 z_1 = d\tau_1 e_0 H_0 z_1 \tau_1/(1-\tau_1)$

c_0+dc_0

c_0

Slope 1-τ_1

Optimal phase-out rate τ_1:
$dM+dW+dB=0$

$\rightarrow \tau_1/(1-\tau_1) = (g_0-1)/e_0$

45°

0

z_1

Earnings z

Figure 7 Optimal bottom marginal tax rate with only intensive labor supply responses. The figure, adapted from Diamond and Saez (2011), depicts the derivation of the optimal marginal tax rate at the bottom in the discrete Mirrlees (1971) model with labor supply responses along the intensive margin only. Let H_0 be the fraction of the population not working. This is a function of $1 - \tau_1$, the net-of-tax rate at the bottom, with elasticity e_0. We consider a small reform around the optimum: The government increases the maximum transfer by c_0 by increasing the phase-out rate by $d\tau_1$ leaving the tax schedule unchanged for those with income above z_1. This creates three effects which cancel out at the optimum. At the optimum, we have $\tau_1/(1 - \tau_1) = (g_0 - 1)/e_0$ or $\tau_1 = (g_0 - 1)/(g_0 - 1 + e_0)$. Under standard redistributive preferences, g_0 is large implying that τ_1 is large.

at the bottom can be written as:

Optimal bottom marginal tax rate in Mirrlees model: $\quad \tau_1 = \dfrac{g_0 - 1}{g_0 - 1 + e_0},$

$$(14)$$

where g_0 is the average social marginal welfare weight on zero earners and $e_0 = -[(1 - \tau_1)/h_0]dh_0/d(1 - \tau_1)$ is the elasticity of the fraction non-working h_0 with respect to the bottom net–of–tax rate $1 - \tau_1$ with a minus sign so that $e_0 > 0$.[84] This formula is proved by Saez (2002a) in the discrete model presented above.[85]

The formula also applies in the standard Mirrlees model although it does not seem to have been ever noticed and formally presented. We present the proof in the standard Mirrlees model in the appendix. In the text, we present a simple graphical proof adapted from Diamond and Saez (2011) using the discrete model with intensive margin responses presented above.

As illustrated on Figure 7, suppose that low ability individuals can choose either to work and earn z_1 or not work and earn zero ($z_0 = 0$). The government offers a transfer $c_0 = -T(0)$ to those not working phased–out at rate τ_1 so that those working receive

---

[84] This elasticity e_0 reflects substitution effects only, as income effects are second order when the marginal tax rate is changed only on a small band of income at the bottom.

[85] It can be obtained from Eq. (13) noting that the average social marginal welfare weight is equal to one so that $\sum_{m\geq 0} (1 - g_m)h_m = 0$. Therefore, $\tau_1/(1 - \tau_1) = (1/e_1)(g_0 - 1)h_0/h_1$. Finally, note that $h_1 e_1 = h_0 e_0$.

on net $c_1 = (1 - \tau_1)z_1 + c_0$. In words, non-workers keep a fraction $1 - \tau_1$ of their
earnings should they work and earn z_1. Therefore, increasing τ_1 discourages some low
income workers from working. Suppose now that the government increases both the c_0
by dc_0 and the phase-out rate by $d\tau_1$ leaving the tax schedule unchanged for those with
income equal to or above z_1 so that $dc_0 = z_1 d\tau_1$ as depicted on Figure 7. The fiscal
cost is $-h_0 dc_0$ but the welfare benefit is $h_0 g_0 dc_0$ where g_0 is the social welfare weight on
non-workers. Because behavioral responses take place along the intensive margin only in
the Mirrlees model, with no income change above z_1, the labor supply of those above z_1
is not affected by the reform. By definition of e_0, a number $dh_0 = d\tau_1 e_0 h_0/(1-\tau_1)$ of low
income workers stop working creating a revenue loss of $-\tau_1 z_1 dh_0 = -dc_0 h_0 e_0 \tau_1/(1-\tau_1)$.
At the optimum, the three effects sum to zero leading to the optimal bottom rate formula
(14). Three points are worth noting about formula (14).

First, if society values redistribution toward zero earners, then g_0 is likely to be large
(relative to 1). In that case, τ_1 is going to be high even if the elasticity e_0 is large. For
example, if $g_0 = 3$ and $e_0 = .5$ then $\tau_1 = 80\%$, a very high phase-out rate. The intuition
is simple: increasing transfers by increasing the phase-out rate is valuable if g_0 is large, the
fiscal cost due to the behavioral response is relatively modest as those dropping out of the
labor force would have had very modest earnings anyway. The phase-out rate is highest
in the Rawlsian case where all the social welfare weight is concentrated at the bottom.[86]

Second and conversely, if society considers that non-workers are primarily free-loaders
taking advantage of transfers, then $g_0 < 1$ is conceivable. In that case, the optimal phase-
out rate is negative and the government provides higher transfers for low income earners
rather than those out-of-work. Naturally, this cannot happen under the standard assump-
tion where social marginal welfare weights decrease with income.

Finally, note that it is not possible to obtain an explicit formula for the optimal
demogrant $-T(0)$ as the demogrant is determined in general equilibrium. This is a
general feature of optimal tax problems (in the optimal linear tax rate, the demogrant was
also deduced from the optimal tax rate τ using the government budget constraint).

5.3.2. Extensive Margin Responses

The optimality of a traditional means-tested transfer program with a high phase-out rate
depends critically on the assumption of intensive labor supply responses. Empirically
however, there is substantial evidence that labor supply responses, particularly among low
income earners, are also substantial along the extensive margin with less compelling evi-
dence of intensive marginal labor supply response.[87] In that case, it is optimal to give

---

[86] In the Rawlsian case, $g_0 = 1/h_0$ and the optimum phase-out rate is almost 100% when the fraction non-working h_0
is small.

[87] Chetty (2012) argues that intensive elasticities are more affected by frictions or inattention issues than extensive
elasticities. This makes it more challenging to identify long-run intensive elasticities. For example, Chetty, Friedman,
and Saez (2012) show that intensive responses to the EITC can also be substantial in the long-run in places where
knowledge about the EITC is high.

higher transfers to low income workers rather than non-workers, which amounts to a negative phase-out rate, as with the current Earned Income Tax Credit (Diamond, 1980; Saez, 2002a).

To see this, consider now a model where behavioral responses of low- and mid-income earners take place through the extensive elasticity only, i.e., whether or not to work, and that earnings when working do not respond to marginal tax rates. Within the general discrete model developed in Section 5.2.2, the extensive model can be obtained by assuming that each individual can only work in one occupation or be unemployed. This can be embodied in the individual utility functions by assuming that $u^i(c_n, n) = -\infty$ for all occupations $n \geq 1$ except the one corresponding to the skill of the individual. This structure implies that the fraction of the population h_n working in occupation n depends only on c_0 and c_n for $n \geq 1$. As a result, and using the fact that $\partial h_n / \partial c_n + \partial h_0 / \partial c_n = 0$, and defining the elasticity of participation $e_n = [(1 - \tau_n)/h_n]dh_n/d(1 - \tau_n)$, Eq. (12) becomes,

Optimal tax rate with extensive responses only: $\dfrac{\tau_n}{1 - \tau_n} = \dfrac{1}{e_n}(1 - g_n).$ (15)

To obtain this result, as depicted on Figure 8, suppose the government starts from a transfer scheme with a positive phase-out rate $\tau_1 > 0$ and introduces an additional small in-work
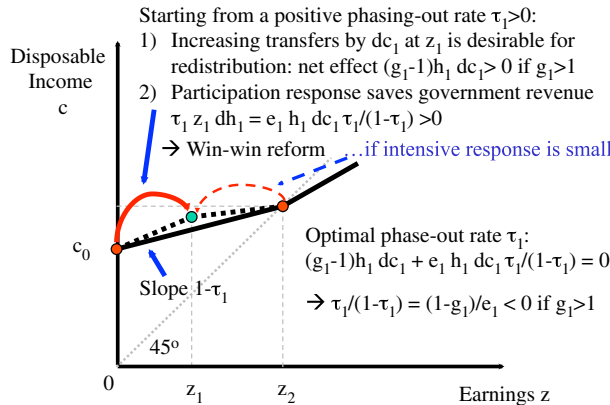

Figure 8 Optimal bottom marginal tax rate with extensive labor supply responses. The figure, adapted from Diamond and Saez (2011), depicts the derivation of the optimal marginal tax rate at the bottom in the discrete model with labor supply responses along the extensive margin only. Starting with a positive phase-out rate $\tau_1 > 0$, the government introduces a small in-work benefit dc_1. Let h_1 be the fraction of low income workers with earnings z_1, and let e_1 be the elasticity of h_1 with respect to the participation net-of-tax rate $1 - \tau_1$. The reform has three standard effects: mechanical fiscal cost $dM = -h_1 dc_1$, social welfare gain, $dW = g_1 h_1 dc_1$, and tax revenue gain due to behavioral responses $dB = \tau_1 z_1 dh_1 = e_1 h_1 dc_1 \tau_1/(1 - \tau_1)$. If $g_1 > 1$, then $dW + dM > 0$. If $\tau_1 > 0$, then $dB > 0$ implying that $\tau_1 > 0$ cannot be optimal. The optimal τ_1 is such that $dM + dW + dB = 0$ implying that $\tau_1/(1 - \tau_1) = (1 - g_1)/e_1$.

benefit dc_1 that increases net transfers to low income workers earning z_1. Let h_1 be the fraction of low income workers with earnings z_1. The reform has again three effects.

First, the reform has a mechanical fiscal cost $dM = -h_1 dc_1$ for the government. Second, it generates a social welfare gain, $dW = g_1 h_1 dc_1$ where g_1 is the marginal social welfare weight on low income workers with earnings z_1. Third, there is a tax revenue gain due to behavioral responses $dB = \tau_1 z_1 dh_1 = e_1[\tau_1/(1-\tau_1)]h_1 dc_1$. If $g_1 > 1$, then $dW + dM > 0$. In that case, if $\tau_1 > 0$, then $dB > 0$, implying that $\tau_1 > 0$ cannot be optimal. The optimal τ_1 is such that

$$0 = dM + dW + dB = h_1 dc_1 \left[g_1 - 1 + e_1 \frac{\tau_1}{1-\tau_1} \right],$$

implying that the optimal phase-out rate at the bottom is given by:

$$\textbf{Optimal bottom tax rate, extensive model: } \tau_1 = \frac{1 - g_1}{1 - g_1 + e_1},$$
$$\tau_1 < 0 \quad \text{if} \quad g_1 > 1, \quad (16)$$

Intuitively, starting with a transfer system with a positive phase-out rate as depicted on Figure 8 and ignoring behavioral responses, an in-work benefit reform depicted on Figure 8 is desirable if the government values redistribution to low income earners. If behavioral responses are solely along the extensive margin, this reform induces some non-workers to start working to take advantage of the in-work benefit. However, because we start from a situation with a positive phase-out rate, this behavioral response increases tax revenue as low income workers still end up receiving a smaller transfer than non-workers. Hence, the in-work benefit increases social welfare implying that a positive phase-out rate cannot be optimal.[88] Another way to see this is the following. Increasing c_0 distorts the labor supply decision of all types of workers who might quit working. In contrast, increasing c_1 distorts labor supply of low-skilled workers only. Hence an in-work benefit is less distortionary than an out-of-work benefit in the pure extensive model.

5.3.3. Policy Practice

In practice, both extensive and intensive elasticities are present. An intensive margin response would induce those earning slightly more than the minimum to reduce labor supply to take advantage of the in-work benefit, thus reducing tax revenue. Therefore,

---

[88] At the optimum, it is always the case that $g_1 < 1 + e_1$ so that the denominator in formula (16) is always positive. To see this, suppose $g_1 \geq 1 + e_1$, then $g_1 - 1 + e_1 \frac{\tau_1}{1-\tau_1} \geq e_1/(1-\tau_1) > 0$ as $\tau_1 < 1$, implying that the reform dc_1 described above is always welfare improving. This result can be understood as follows. Suppose we start from an initial tax system (not optimal) where $g_1 > 1 + e_1$, i.e., low-skilled workers are deserving and their elasticity e_1 is not too high. In such a configuration, it is always desirable to increase in-work benefits for low-skilled workers. Increasing in-work benefits reduces g_1 as low-skilled workers become less and less in need of additional support. At the optimum where (16) holds, $g_1 < 1 + e_1$. In the extreme case with no behavioral responses, τ_1 should be set so that $g_1 = 1$. Conversely, when the elasticity e_1 is very large, the optimal bottom tax rate goes to zero.

the government has to trade off the two effects. If, as empirical studies show (see e.g., Blundell & MaCurdy, 1999 for a survey), the extensive elasticity of choosing whether to participate in the labor market is large relative to the intensive elasticity of choosing how many hours to work, initially low (or even negative) phase-out rates combined with high positive phase-out rates further up the distribution would be the optimal profile.

In recent decades in most OECD countries, a concern arose that traditional welfare programs overly discouraged work and there has been a marked shift toward lowering the marginal tax rate for low earners through a combination of: (a) introduction and then expansion of in-work benefits such as the Earned Income Tax Credit in the United States or the Family Credit in the United Kingdom;[89] (b) reduction of the statutory phase-out rates in transfer programs for earned income as under the U.S. welfare reform; and (c) reduction of payroll taxes for low income earners.[90] Those reforms are consistent with the logic of the optimal tax model we have outlined, as they both encourage labor force participation and provide transfers to low income workers seen as a deserving group. As we saw on Figure 2, the current US system imposes marginal tax rates close to zero on the first $15,000 of earnings but significantly higher marginal rates between $15,000 and $30,000.

How can we explain however that means-tested social welfare programs with high phase-out rates were widely used in prior decades? Historically, most means-tested transfer programs started as narrow programs targeting specific groups deemed unable to earn enough such as widows with children, the elderly, or the disabled. For example, the ancestor of the traditional US welfare program (Aid for Families with Dependent Children, renamed Temporary Aid for Needy Families after the 1996 welfare reform) were "mothers' pensions" state programs providing help primarily to widows with children and no resources (Katz, 1996). If beneficiaries cannot work but differ in terms of unearned income (for example, the presence of a private pension), then the optimal redistribution scheme is indeed a transfer combined with a 100% phasing-out rate. As governments expanded the scope of transfers, a larger fraction of beneficiaries were potentially able to work. The actual tax policy response to this moral hazard problem over the last few decades has been remarkably close to the lessons from optimal tax theory we have outlined.

Note that following the Reagan and Thatcher conservative revolutions two other elements likely played a role in the shift from traditional means-tested programs toward in-work benefits. First, it is conceivable that society has less tolerance for non-workers living off government transfers because it believes, rightly or wrongly, that most of such non-workers could actually work and earn a living on their own absent government transfers. This means that the social welfare weights on non-workers has fallen relative to the social welfare weights on workers, and especially low income workers. This effect can be captured in our model simply assuming that social welfare weights change (see Section 7 for a discussion of how social welfare weights could be formed in non-utilitarian

---

[89] See OECD (2005, chap. 3) for a review of all the in-work benefits introduced in OECD countries up to year 2004.
[90] See OECD 2011b for a summary of such payroll tax reductions in OECD countries.

contexts). Second and related, the perception that relying on transfers generates negative externalities on children or neighbors through a "culture of welfare dependency" might have increased. Such externalities are not incorporated in our basic model but could conceivably be added. In both cases, perceptions of the public and actual facts do not necessarily align (see e.g., Bane & Ellwood, 1994 for a detailed empirical analysis).

6. EXTENSIONS

6.1. Tagging

We have assumed that $T(z)$ depends only on earnings z. In reality, the government can observe many other characteristics (denoted by vector X) also correlated with ability (and hence social welfare weights), such as gender, race, age, disability, family structure, height, etc. Hence, the government could set $T(z, X)$ and use the characteristic X as a "tag" in the tax system. There are two noteworthy theoretical results.

First, if characteristic X is *immutable* then there should be full redistribution across groups with different X. This can be seen as follows. Suppose X is a binary 0–1 variable. If the average social marginal welfare weight for group 1 is higher than for group 0, a lump sum tax on group 0 funding a lump sum transfer on group 1 will increase total social welfare.

Second, if characteristic X is not *immutable*, i.e., it can be manipulated through cheating,[91] then it is still desirable to make taxes depend on X (in addition to z). At the optimum however, the redistribution across the X groups will not be complete. To see this, suppose again that X is a binary 0–1 variable and that we start from a pure income tax $T(z)$. As X is correlated with ability, the average social marginal welfare weight for group 1 is different from the one for group 0. Let us assume it is higher. In that case, a small lump sum transfer from group 0 to group 1 increases social welfare, absent any behavioral response. As X is no longer immutable, this small transfer might induce some individuals to switch from group 0 to group 1. However, because we start from a unified tax system, at the margin those who switch do not create any first order fiscal cost (nor any welfare cost through the standard envelope theorem argument).[92]

Those points on tagging have been well known in the literature for decades following the analysis of Akerlof (1978) and Nichols and Zeckhauser (1982) for tagging disadvantaged groups for welfare benefits. It has received recent attention in Mankiw and Weinzierl (2010) and Weinzierl (2011) who use the examples of height and age respectively to argue

---

[91] A good example would be disability status that can only be imperfectly observed and that individuals can fake to some extent.

[92] Note that this derivation assumes that labor supply choices z are independent of X. This assumption is reasonable when X is manipulated through cheating only but would not necessarily hold if X was manipulated through real choices (e.g., hurting oneself to becoming truly disabled).

that the standard utilitarian maximization framework fails to incorporate important elements of real tax policy design.

Indeed, in reality, actual tax systems depend on a very limited set of characteristics besides income. Those characteristics are primarily family structure (in particular the number of dependent children), disability status (for permanent and temporary disability programs). Hence, characteristics used reflect direct "need" (for example, the size of the household relative to income), or direct "ability-to-earn" (as is the case with disability status). To the best of our knowledge, the case for using indirect tags correlated with ability in the tax or transfer system has never been made in practice in the policy debate, implying that society does have a strong aversion for using indirect tags. We come back to this issue in Section 7 when we discuss the limits of utilitarianism.

6.2. Supplementary Commodity Taxation

The government can also implement differentiated commodity taxation in addition to nonlinear income taxes and transfers. The usual hypothesis is that commodity taxes have to be linear because of retrading (see e.g., Guesnerie, 1995, Chapter 1). The most common form of commodity taxation, value added taxes and general sales taxes, do display some variation in rates across goods, with exemptions for specific goods, such as food or housing. Such exemptions are in general justified on redistributive grounds. The government also imposes additional taxes on specific goods such as gasoline, tobacco, alcohol, airplane tickets, or motor vehicles.[93] Here, we want to analyze whether it is desirable to supplement the optimal nonlinear labor income tax with differentiated linear commodity taxation.

Consider a model with K consumption goods $c = (c_1, \ldots, c_K)$ with pre-tax prices $p = (p_1, \ldots, p_K)$. Individual i derives utility from the K consumption goods and earnings supply according to a utility function $u^i(c_1, \ldots, c_K, z)$. The question we want to address is whether the government can increase social welfare using differentiated commodity taxation $t = (t_1, \ldots, t_K)$ in addition to nonlinear optimal income tax on earnings z. Naturally, adding fiscal tools cannot reduce social welfare. However, Atkinson and Stiglitz (1976) demonstrated the following.

Atkinson–Stiglitz Theorem. Commodity taxes cannot increase social welfare if utility functions are weakly separable in consumption goods vs. leisure and the subutility of consumption goods is the same across individuals, i.e., $u^i(c_1, \ldots, c_K, z) = U^i(v(c_1, \ldots, c_K), z)$ with the subutility function $v(c_1, \ldots, c_K)$ homogenous across individuals.

The original proof by Atkinson and Stiglitz (1976) was based on optimum conditions and not intuitive. Recently, Laroque (2005) and Kaplow (2006) have simultaneously and independently proposed a much simpler and intuitive proof that we present here.

---

[93] Traditionally, excise taxes have been used on goods where transactions were relatively easy for the government to monitor. In modern times, current excise taxes are often justified because of externalities (e.g., gasoline taxes because of pollution or global warming), or "internalities" (e.g., tobacco and addiction in models with self-control issues). We assume away such effects in what follows. Externalities are covered in the handbook chapter by Bovenberg and Goulder (2002).

Proof. The idea of the proof is that a tax system $(T(\cdot), t)$ that includes both a nonlinear income tax and a vector of commodity taxes can be replaced by a pure income tax $(\overline{T}(\cdot), t = 0)$ that keeps all individual utilities constant and raises at least as much tax revenue.

Let $V(p + t, y) = \max_c v(c_1, \ldots, c_K)$ subject to $(p + t) \cdot c \leq y$ be the indirect utility of consumption goods common to all individuals. Consider replacing $(T(\cdot), t)$ with $(\overline{T}(\cdot), t = 0)$ where $\overline{T}(z)$ is defined such that $V(p + t, z - T(z)) = V(p, z - \overline{T}(z))$. Such a $\overline{T}(z)$ naturally exists (and is unique) as $V(p, y)$ is strictly increasing in y. This implies that $U^i(V(p + t, z - T(z)), z) = U^i(V(p, z - \overline{T}(z)), z)$ for all z. Hence, both the utility and the labor supply choice are unchanged for each individual i.

By definition of an indirect utility, attaining utility of consumption $V(p, z - \overline{T}(z))$ at price p costs at least $z - \overline{T}(z)$. Let c^i be the consumer choice of individual i under the initial tax system $(T(\cdot), t)$. Individual i attains utility $V(p, z - \overline{T}(z)) = V(p + t, z - T(z))$ when choosing c^i. Hence $p \cdot c^i \geq z - \overline{T}(z)$. As $(p + t) \cdot c^i = z - T(z)$, we have $\overline{T}(z) \geq T(z) + t \cdot c^i$, i.e., the government collects more taxes with $(\overline{T}(\cdot), t = 0)$ which completes the proof. QED.

Intuitively, with separability and homogeneity, conditional on earnings z, the consumption choices $c = (c_1, \ldots, c_K)$ do not provide any information on ability. Hence, differentiated commodity taxes $t_1, \ldots, t_K$ create a tax distortion with no benefit and it is better to do all the redistribution with the individual nonlinear income tax. With the weaker linear income taxation tool, stronger assumptions on preferences, namely linear Engel curves uniform across individuals, are needed to obtain the commodity tax result (Deaton 1979).[94] Intuitively, in the linear tax case, unless Engel curves are linear, commodity taxation can be useful to "non-linearize" the tax system.

Heterogeneous Preferences. Saez (2002b) shows that the Atkinson–Stiglitz theorem can be naturally generalized to cases with heterogeneous preferences. No tax on commodity k is desirable under three assumptions: (a) conditional on income z, social marginal welfare weights are uncorrelated with the levels of consumption of good k, (b) conditional on income z, the behavioral elasticities of earnings are uncorrelated with the consumption of good k, and (c) at any income level z, the average individual variation in consumption of good k with z is identical to the cross-sectional variation in consumption of good k with z.

Assumption (a) is clearly necessary and might fail when earnings z is no longer a sufficient statistic for measuring welfare. For example, if some individuals face high

---

[94] The Laroque-Kaplow method can be easily adapted to the linear earnings tax case. Consider a linear earnings tax with tax rate τ and demogrant R. The same proof carries over if any tax system (τ, R, t) can be replaced by a pure income tax $(\bar{\tau}, \bar{R}, t = 0)$ such that $V((1 - \tau)z + E, p + t) = V((1 - \bar{\tau})z + \bar{E}, p)$ for all z. This is possible if and only if $V(y, p)$ takes the linear form $\phi(p) \cdot y + \psi(p)$ (up to an increasing transformation). This in turn is equivalent to having a direct subutility of consumption of the form $v(c_1 - c_1^0(q), \ldots, c_K - c_K^0(q))$ homogeneous of degree 1 (up to an increasing transformation) which delivers affine Engel curves of the form $c_k(y, q) = c_k^0(q) + d_k(q)y$. Importantly, the subutility has to be uniform across individuals.

uninsured medical expenses due to poor health, then this assumption would not hold, and it would be desirable to subsidize health expenditures.[95] However, when heterogeneity in consumption reflects heterogeneity in preferences and not in need, assumption (a) is a natural assumption.

Assumption (b) is a technical assumption required to ensure that consumption of specific goods is not a tag for low responsiveness of labor supply to taxation. For example, if consumers of luxury cars happened to have much lower labor supply elasticities than average, it would become efficient to tax luxury cars as a way to indirectly tax more the earnings of those less responsive individuals. In practice, too little is known about the heterogeneity in labor supply across individuals to exploit such possibilities. Hence, assumption (b) is also a natural assumption.

Assumption (c) is the critical assumption. When it fails, the thought experiment to decide on whether commodity k ought to be taxed is the following. Suppose high ability individuals are forced to work less and earn only as much as lower ability individuals. In that scenario, if higher ability individuals consume more of good k than lower ability individuals, then taxing good k is desirable. This can happen for two reasons. First, high ability people may have a relatively higher taste for good k (independently of income) in which case taxing good k is a form of indirect tagging of high ability. Second, good k is positively related to leisure, i.e., consumption of good k increases when leisure increases keeping after-tax income constant. This suggests taxing more holiday-related expenses and subsidizing work-related expenses such as child care.

In general the Atkinson–Stiglitz assumption is a good starting place for most goods. This implies that lower or zero VAT rates on some goods for redistribution purposes is inefficient (in addition to being administratively burdensome). Under those assumptions, eliminating such preferential rates and replacing them with a more redistributive income tax and transfer system would increase social welfare.[96]

6.3. In-Kind Transfers

As we discussed in Section 3, the largest transfer programs are in-kind rather than cash. OECD countries in general provide universal public health care benefits and public education. They also often provide in-kind housing or nutrition benefits on a means-tested basis.

As is well known, from a rational individual perspective, if the in-kind benefit is tradable, it is equivalent to cash. Most in-kind benefits however are not tradable. In that

---

[95] It also fails in the case with bequests as earnings are no longer a sufficient statistic for lifetime resources in that case. This implies that positive bequest taxes are desirable when the redistributive tastes of the government are strong enough (Piketty and Saez (2012a,2012b)).

[96] This is one of the main recommendations of the recent Mirrlees review (Mirrlees, 2011). The political issue is that it would be difficult in practice to ensure that the VAT reform would indeed by accompanied by truly compensating changes on the income tax and transfer side. Boadway (2012) provides a comprehensive summary of the discussions and applications of the Atkinson and Stiglitz theorem in the literature.

case, recipients may be forced to overconsume the good provided in-kind and would instead prefer to receive the cash equivalent value of the in-kind transfer. Therefore, from a narrow rational individual perspective, cash transfers dominate in-kind transfers. From a social perspective, three broad lines of justification have been provided in favor of in-kind benefits.[97]

1. *Commodity Egalitarianism:* A number of goods, such as education or health care are seen as rights everybody in society is entitled to.[98] Those goods are hence put in the same category as other rights that democratic governments offer to all citizens without distinction such as protection under the law, free speech, right to vote, etc. The difficulty with this view is that it does not say which level of education or health care should be seen as a right.

2. *Paternalism:* The government might want to impose its preferences on transfer recipients. For example, voters might support providing free shelter and free meals to the homeless but would oppose giving them cash that might be used for alcohol or tobacco consumption. In that case, recipients would rather get the cash equivalent value of the non-cash transfers they get but society's paternalistic views prevail upon recipients' preferences. Those arguments have been developed mostly by libertarians to criticize in-kind benefits (e.g., Milton Friedman was favorable to basic redistribution through a negative income tax cash transfer rather than in-kind benefits).

3. *Individual Failures:* Related, recipients could themselves realize that, if provided with only cash, they might choose too little health care, education, or retirement savings for their long-term well being, perhaps because of lack of information or self-control problems (e.g., hyperbolic discounting is an elegant way to model such self-control issues). In this case, recipients understand that non-cash benefits are in their best interest. Hence, recipients would actually support getting such non-cash benefits instead of the equivalent cash value. This type of rationalization for non-cash transfers hence differs drastically from the paternalistic view. The fact that all advanced economies systematically provide large amounts of non-cash benefits universally (retirement, health, education) through a democratic process is more consistent with the "individual failures" scenario than the "paternalism" scenario. The case of education, and especially primary education, is particularly important. Children cannot be expected to have fully forward looking rational preferences. Parents make educational choices on behalf of their children and most—but not all—parents have the best interests of their children at heart. Compulsory and free public education is a simple way for the government to ensure that all children get a minimum level of education regardless of how caring their parents are.

[97] The traditional externality and public good justification, analyzed extensively, may also apply to some although not all types of non-cash benefits and is left aside here.

[98] Retirement benefits, although not strictly speaking in-kind benefits, can also be seen as non-cash benefits because they are not transferrable over time, i.e., a young worker typically cannot borrow against her future retirement benefits.

4. *Second-best Efficiency:* A number of studies have shown that, with limited information and limited policy tools, non-cash benefits can actually be desirable in a "second-best" equilibrium. In-kind benefits can be used by the government to relax the incentive constraint created by the optimal tax problem. This point was first noted by Nichols and Zeckhauser (1982) and later developed in a number of studies (see Currie & Gahvari, 2008 and Boadway, 2012, Chapter 4 for detailed surveys). Those results are closely related to the Atkinson and Stiglitz (1976) theorem presented above. If the utility function is not separable between consumption goods and leisure, then we know that commodity taxation is useful to supplement optimal nonlinear earnings taxation. By the same token, it can be shown that providing an in-kind transfer of a good complementary with work is desirable because it makes it relatively more costly for high skill people to work less. Although such "second-best" arguments have attracted the most attention in the optimal tax literature, they are second order in the public debate which focuses primarily on the other justifications we discussed above.

6.4. Family Taxation

In practice, the treatment of families raises important issues. Any tax and transfer system must make a choice on how to treat singles vs. married households and how to make taxes and transfers depend on the number of children. There is relatively little normative work on those questions, in large part because the standard utilitarian framework is not successful at capturing the key trade offs. Kaplow (2008), Chapter 8 provides a detailed review.

Couples. Any income tax system needs to decide how to treat couples vs. single individuals. As couples typically share resources, welfare is best measured by family income rather than individual income. There are two main treatments of the family in actual tax (or transfer) systems. (a) The individual system where every person is taxed separately based on her individual income. In that case, couples are treated as two separate individuals. As a result, an individual system does not impose any tax or subsidy on marriage as tax liability is independent of living arrangements. At the same time, it taxes in the same way a person married to a wealthy spouse vs. a person married to a spouse with no income. (b) The family system where the income tax is based on total family income, i.e., the sum of the income of both spouses in case of married couples. The family system can naturally modulate the tax burden based on total family resources, which best measures welfare under complete sharing within families. However and as a result, a family tax system with progressive tax brackets cannot be neutral with respect to living arrangements, creating either a marriage tax or a marriage subsidy. Under progressive taxation, if the tax brackets for married couples are the same as for individuals, the family system typically creates a marriage tax. If the tax brackets for married couple are twice as wide as for individuals, the family system typically creates a marriage subsidy.[99]

---

[99] The US system creates marriage subsidies for low to middle income families and marriage taxes for high income families with two earners.

Hence and as is well known, it is impossible to have a tax system that simultaneously meets three desirable properties: (1) the tax burden is based on family income, (2) the tax system is marriage neutral, and (3) the tax system is progressive (i.e., the tax system is not strictly linear). Although those properties clearly matter in the public debate, it is not possible to formalize their trade off within the traditional utilitarian framework as the utilitarian principle cannot put a weight on the marriage neutrality principle.

If marriage responds strongly to any tax penalty or subsidy, it is better to reduce the marriage penalty/subsidy and move toward an individualized system. This issue might be particularly important in countries (such as Scandinavian countries for example), where many couples cohabit without being formally married and as it is difficult (and intrusive) for the government to observe (and monitor) cohabitation status.

Traditionally, the labor supply of secondary earners—typically married women—has been found to be more elastic than the labor supply of primary earners—typically married men (see Blundell & MaCurdy, 1999 for a survey). Under the standard Ramsey taxation logic, this implies that it is more efficient to tax secondary earners less (Boskin & Sheshinski, 1983). If the tax system is progressive, this goal is naturally achieved under an individual-based system as secondary earners are taxed on their sole earnings. Note however that the difference in labor supply elasticities between primary and secondary earners has likely declined over time as more and more married women work (Blau & Kahn, 2007).

In practice, most OECD countries have switched from family based to individual-based income taxation. In contrast, transfer systems remain based on family income. It is therefore acceptable to the public that a spouse with modest earnings would face a low tax rate, no matter how high the earnings of her/his spouse are.[100] In contrast, it appears unacceptable to the public that a spouse with modest earnings should receive means-tested transfers if the earnings of his or her spouse are high. A potential explanation could be framing effects as direct transfers might be more salient than an equivalent reduction in taxes. Kleven, Kreiner, and Saez (2009b) offer a potential explanation in a standard utilitarian model with labor supply where they show that the optimal joint tax system is to have transfers for non-working spouses (or equivalently taxes on secondary earnings) that *decrease* with primary earnings. The intuition is the following. With concave utilities, the presence of secondary earnings make a bigger difference in welfare when primary earnings are low than when primary earnings are large. Hence, it is more valuable to compensate one earner couples (relative to two earner couples) when primary earnings are low. This translates into an implicit tax on secondary earnings that decreases with primary earnings. Such negative jointness in the tax system is approximately achieved by having family based means-tested transfers along with individually based income taxation.

---

[100] Note that under a progressive and individual based tax system, only small earnings of secondary earners face low tax rates. As secondary earnings increase, they get taxed at progressively higher rates.

Children. Most tax and transfer systems offer tax reductions for children or increases in benefits for children. The rationale for such transfers is simply that, conditional on income z, families with more children are more in need of transfers and have less ability to pay taxes. The interesting question that arises is how the net transfer (additional child benefits or reduction in taxes) per additional child should vary with income z. On the one hand, the need for children related transfers is highest for families with very small incomes. On the other hand, the cost of children is higher for families with higher incomes particularly when parents work and need to purchase childcare.

Actual tax and transfers do seem to take both considerations into account. Means-tested transfers tend to offer child benefits that are phased-out with earnings. Income taxes tend to offer child benefits that increase with income for two reasons. First, the lowest income earners do not have taxable income and hence do not benefit from child-related tax reductions. Second, child-related tax reductions are typically a fixed deduction from taxable income which is more valuable in upper income tax brackets. Hence, the level of child benefits tends to be U-shaped as a function of earnings. Two important qualifications should be made.

First, as mentioned in Section 5.3.3, a number of countries have introduced in-work benefits that are tied to work and presence of children. This tends to make child benefits less decreasing with income at the low income end. In the United States, because of the large EITC and child tax credits and small traditional means-tested transfers, the benefit per child is actually increasing with family earnings at the bottom. Second, another large child benefit often subsidized or government provided is pre-school child care (infant child care, kindergarten starting at age 2 or 3, etc.). Such child care benefits are quantitatively large and most valuable when both parents work or for single working parents. Hence, economically, they are a form of in-kind in-work benefit which also promotes labor force participation (see OECD, 2006, chap. 4, Figure 4.1, p.129 for an empirical analysis). It is perhaps not a coincidence that cash in-work benefits for children are highest in the US and the UK, countries which provide minimal child care public benefits. Understanding in that context whether a cash transfer or an in-kind child care benefit is preferable is an interesting research question that has received little attention.

Child-related benefits raise two additional interesting issues.

First, families do not take decisions as a single unit (Chiappori, 1988). Interestingly, in the case of children, cash transfers to mothers (or grandmothers) have larger impacts on children's consumption than transfers to fathers. This has been shown in the UK context (Lundberg, Pollak, & Wales, 1997) when the administration of child tax benefits was changed from a reduction in tax withholdings of parents (often the father) to a direct check to the mother. Similar effects have been documented in the case of cash benefits for the elderly in South Africa (Duflo, 2003). This evidence suggests that in-kind benefits (such as child care or pre-school) might be preferable if the goal is to ensure that resources go toward children. As mentioned above, primary education is again the most important

example of in-kind benefits designed so that children benefit regardless of how caring parents are.

Second, child benefits might promote fertility. A large empirical literature has found that child benefits have sometimes positive but in general quite modest effects on fertility (see Gauthier, 2007 for a survey). There can be externalities (both positive and negative) associated with children. For example, there can be congestion effects (such as global warming) associated with larger populations. Alternatively, declines in populations can have adverse effects on sustainability of pay-as-you-go pension arrangements. Such externalities should be factored into discussions of optimal child benefits.

6.5. Relative Income Concerns

Economists have long been interested in the possibility that individuals care not only about their absolute income but also their income relative to others. Recently, substantial evidence coming from observational studies (e.g., Luttmer, 2005), lab experiments (e.g., Fehr & Schmidt, 1999), and field experiments (Card, Mas, Moretti, & Saez 2012), provide support for relative income effects. A number of optimal tax studies have incorporated relative income in the analysis (Boskin & Sheshinski, 1978 analyze the linear income tax case and Oswald, 1983 and Tuomala, 1990, Chapter 8 consider the nonlinear income tax case). Those studies find that in general relative income concerns tend to increase optimal tax rates. Relative income effects can be modeled in a number of ways. The simplest way, which we consider here, is to posit that individual utility also depends on the utility of others.[101]

Relative income concerns affect optimal tax analysis in two ways. First, it changes the social marginal welfare weights as a decrease in the utility of others has a direct effect on one's utility (keeping one's work and income situation constant), creating externalities. In our view, the simplest way to capture this effect is to consider that those externalities affect the social welfare weights. If a decrease in a person's income increases others' utility, then the social welfare weight on this person ought to be reduced by this external effect. Whether such externalities should be factored in the social welfare function is a deep and difficult question. Surely, hurting somebody with higher taxes for the sole satisfaction of envy seems morally wrong, Hence, social welfare weights should not be allowed to be negative for anybody no matter how strong the envy effects. At the same, it seems to us that relative income concerns are a much more powerful and realistic way to justify social welfare weights decreasing with income than standard utilitarianism with concave utility of consumption.

Second, relative income concerns affect labor supply decisions. For example, if utility functions are such that $u(c/\bar{c}, z)$ with $\bar{c}$ average consumption in the economy, then a proportional tax on consumption affects c and $\bar{c}$ equally and hence has no impact on

---

[101] Alternatives could be to make individual utility depends on the earnings or consumption of others.

labor supply. This might be a simple explanation for why labor supply is relatively inelastic with respect to secular increases in wage rates over the long-term process of economic growth (Ramey and Francis, 2009).[102] This labor supply channel effect is fully captured by the behavioral response elasticity and hence does not change the optimal tax formulas.

As an illustration, let us go back to the optimal top tax rate analysis from Section 5.1 with a small variation $d\tau$ in the top tax rate. The key difference in the analysis is that the reduction in welfare for top bracket earners would now have a positive externality on the utility of lower income individuals. As long as this external effect is weakly separable from labor supply choices, i.e., $U^i(u^i(c, z), \bar{u}_{-i})$ where $u^i(c, z)$ is the standard utility function and $\bar{u}_{-i}$ is the vector of utilities of all other (non i) individuals, the individual earnings z^i decisions are not affected by the external effect. The external effect is proportional to the direct welfare effect on top bracket earners and the strength of the externality. Therefore, the external effect simply reduces the social marginal value of consumption of top bracket earners from g to $\hat{g}$. The optimal tax formula retains the same form as before $\tau = (1 - \hat{g})/(1 - \hat{g} + a \cdot e)$.

In sum, we think that relative income concerns are a useful way to interpret and justify optimal tax analysis and can be incorporated within standard optimal tax analysis.

6.6. Other Extensions

Endogenous Wages. The standard assumption in optimal labor income tax theory is that pre-tax wage rates are exogenous, i.e., that there is perfect substitutability between skills in production. Interestingly, in the discrete occupational models we have introduced in Section 5.2.2, this assumption can be relaxed without affecting the general optimal tax formula (12). To see this, consider a general production function $F(h_1, \ldots, h_N)$ of the consumption good with constant returns to scale.[103] In that case, wages are set by marginal product $z_n = \partial F/\partial h_n$. The maximization of the government can be rewritten as choosing $(c_0, \ldots, c_N)$ to maximize

$$SWF = \int_i \omega_i G(u^i(c_n(i), n(i)))d\nu(i) \quad \text{s.t.} \quad \sum_n h_n c_n + E \leq F(h_1 \cdots h_N)(p).$$

Note that any explicit reference to wages z_n has disappeared from this maximization problem and the first order condition with respect to c_n immediately leads to the same optimal tax formula (12).

The intuition in a basic two skill model is the following. Suppose an increase in high skill taxes leads to a reduction in high skill labor supply and hence an increase in high skill wages (and a decrease in low skill wages) through demand effects. Because of the

---

[102] An alternative explanation is that income and substitution effects cancel out so that large uncompensated increases in wage rates have little effect on labor supply.

[103] If returns were not constant, there would be pure profits, the results would carry through assuming that pure profits can be taxed 100%.

absence of profits, those demand effects are a pure transfer from low to high skill workers. Therefore, the government can readjust the tax on high and low skills to offset those demand effects on the net consumption levels at no net fiscal cost, leaving the optimal tax formula unchanged.[104]

Theoretically, this result arises because the discrete occupational model is effectively mathematically identical to a Diamond and Mirrlees (1971), optimal commodity tax model where each occupation is a specific good taxed at a specific rate. As is well known from Diamond and Mirrlees (1971), optimal Ramsey tax formulas depend solely on consumers' demand and do not depend on production functions. This generates two important additional consequences. First, the production efficiency result of Diamond and Mirrlees (1971) carries over to the discrete occupational choice model, implying that distortions in the production process or tariffs (in the case of an open economy) are not desirable. Second, in an extended model with many consumption goods, the theorem of Atkinson and Stiglitz (1976) also carries over to the discrete occupational choice model. Namely, differentiated commodity taxation is not desirable to supplement optimal nonlinear earnings taxation under the standard separability assumption presented above. Those results are formally proven in Saez (2004b). They stand in sharp contrast to results obtained in the Stiglitz (1982) discrete model with endogenous wages where it is shown that the optimal tax formulas are affected by endogenous wages (Stiglitz, 1982), and where the production efficiency theorem and the Atkinson–Stiglitz theorem do not carry over (Naito, 1999). Saez (2004b) argues that the occupational model best captures the long-term when individuals choose their occupations while the Stiglitz (1982) model captures a short-term situation where individuals have fixed skills and only adjust hours of work.

Workfare, Take–Up Costs, and Screening. Workfare can be defined as requiring transfer beneficiaries to work, typically for a public project. In its extreme form, the work required has no productive value. In that case, workfare is similar to imposing an ordeal, such as time consuming take-up costs, on welfare beneficiaries. The literature has focused primarily on such "useless workfare requirements." Besley and Coate (1992) show that, if the government cares about poverty measured by net-income rather than individual utilities, it can be optimal to impose workfare. In their model, workfare screens away higher wage individuals who have a higher opportunity cost of time.[105]

Cuff (2000) shows, in a standard Stiglitz (1982) two-type discrete model that a useless workfare program is never desirable with a standard welfarist objective. Interestingly, Cuff

---

[104] The same result applies when considering differentiated linear taxation of capital and labor income. What matters for optimal tax formulas are the supply elasticities of labor (and capital) and the effects on the prices of factors are again irrelevant. Taxing labor more reduces labor supply, increases the wage rate, and reduces the return on capital, creating indirect redistribution from capital earners to labor earners. However, this indirect redistribution is irrelevant for optimal tax analysis as the government can adjust the capital and labor tax rates to fully offset it at no fiscal cost.

[105] Related, Kleven and Kopczuk (2011) show that imposing complex take-up rules that improve screening but reduce take-up is optimal when the government objective is poverty alleviation instead of standard welfare.

(2000) then extends the analysis to include heterogeneity in tastes for work (in addition to the standard wage rate heterogeneity). When there are lazy vs. hard working low skill workers and when society does not like to redistribute toward lazy low skill workers, workfare can become desirable. This is because work requirements are more costly to lazy types than hard working types.

In practice, finding ordeals which hurt more the undeserving beneficiaries than the deserving beneficiaries seems difficult. In particular, if society feels that welfare is too generous, it is more efficient to cut benefits directly rather than impose ordeals. Both reduce welfare benefits (and hence the incentives to become a recipient), but at least direct cuts save on government spending.

Screening mechanisms that also impose costs on recipients, (e.g., filing out forms, medical tests, etc.) can be desirable when they are successful in screening deserving recipients (e.g., the truly disabled) vs. undeserving recipients (e.g., those faking disability). Diamond and Sheshinski (1995) propose an analysis along those lines in the case of disability insurance (see also the chapter by Chetty and Finkelstein in this volume for more details on optimal social insurance). The key difference with useless workfare or ordeals is that such screening is directly designed at separating deserving vs. undeserving recipients. It is very unlikely that blanket ordeals can achieve this. Today, data driven screening (i.e., checking administrative databases for potential earnings, etc.) are far more powerful and efficient than direct in person screening (and a lot less intrusive for recipients).

Minimum Wages. The minimum wage is another policy tool that can be used for redistribution toward low skill workers. At the same time minimum wages can create unemployment among low skill workers, creating a trade off between equity and efficiency. A small literature has examined the desirability of minimum wages in addition to optimal taxes and transfers in the standard competitive labor market with endogenous wage rates (as in the model discussed above).[106]

Lee and Saez (2012) use the occupational model of Section 5.3.2 with endogenous wages and prove two results. First, they show that a binding minimum wage is desirable under the strong assumption that unemployment induced by the minimum wage hits the lowest surplus workers first. The intuition for this result is simple and can be understood using Figure 8. Suppose a minimum wage is set at level z_1 and that transfers to low-skilled workers earning z_1 are increased. The presence of the minimum wage at z_1 rations low skill work and effectively prevents the labor supply responses from taking place. Some non-workers would like to work and earn z_1 but cannot find jobs because those jobs are rationed by the minimum wage. Therefore, the minimum wage enhances the ability of the government to redistribute (via an EITC type benefit) toward low skill workers.

Second, when labor supply responses are along the extensive margin only, which is the empirically relevant case, the co-existence of a minimum wage with a positive tax rate

[106] A larger literature has considered minimum wages in labor markets with imperfections that we do not review here.

on low-skilled work is always (second-best) Pareto inefficient. A Pareto improving policy consists of reducing the pre-tax minimum wage while keeping constant the post-tax minimum wage by increasing transfers to low-skilled workers, and financing this reform by increasing taxes on higher paid workers. Importantly, this result is true whether or not rationing induced by the minimum wage is efficient or not. This result can also rationalize policies adopted in many OECD countries in recent decades that have decreased the minimum wage while reducing the implicit tax on low skill work through a combination of reduced payroll taxes for low skill workers and in-work benefits of the EITC type for low skill workers.

Optimal Transfers in Recessions. In practice, some transfers (such as unemployment insurance in the United States) can be made more generous during recessions. Traditionally, optimal policy over the business cycle has been analyzed in the macro-economics literature rather than the public economics literature.[107] The macro-economics literature, however, rarely focuses on distributional issues. There are three channels through which recessions can affect the calculus of optimal transfers for those out-of-work.

First, recessions are a time of high unemployment where people want to work but cannot find jobs. This suggests that employment is limited by demand effects rather than the supply effects of the traditional optimal tax analysis. As a result, in recessions, unemployment is likely to be less sensitive to supply-side changes in search efforts and job search is likely to generate a negative externality on other job seekers in the queue. Landais, Michaillat, and Saez (2010) capture this effect in a search model where job rationing arises in recessions and show that unemployment insurance should be more generous during recessions. Crépon, Esther, Marc, Roland and Philippe (in press), using a large scale job placement aid randomized experiment in France, show that indeed there are negative externalities of job placement aid on other job seekers and that those externalities are larger when unemployment is high.

Second, in recessions, the ability to smooth consumption might be reduced, as the long-term unemployed might exhaust their buffer stock savings and might face credit constraints. This implies that the gap in social marginal utility of consumption between workers and non-workers might grow during recessions, further increasing the value of redistributing from workers to the unemployed (Chetty, 2008).

Third and related, individuals are less likely to be responsible for their unemployment status in a recession than in an expansion. In an expansion when jobs are easy to find, long unemployment spells are more likely to be due to low search efforts than in a recession when jobs are difficult to find even with large search efforts. If society wants to redistributive toward the hard-searching unemployed—i.e., those who would not have found jobs even absent unemployment benefits—then it seems desirable to have time

---

[107] Stabilization policy was one of the three pillars of public policy in the famous Musgrave terminology, the other two being the allocative and redistributive policies.

limited benefits during good times combined with expanded benefit durations in bad times. We will come back to such non-utilitarian social preferences in Section 7.

Education Policy. Education plays a critical role in generating labor market skills. All advanced economies provide free public education at the K–12 level and heavily subsidize higher education. As we have seen earlier, there is a strong rationale for providing K–12 public education to correct potential parenting failures. For higher education, the presence of credit constraints might lead to suboptimal educational levels, providing a strong rationale for government provision of loans (see e.g., Lochner and Monge, 2011).[108] However, governments in advanced economies not only provide loans but also direct subsidies to higher education. Direct subsidies could be justified by "behavioral considerations" if a significant fraction of young adults are not able to make wise educational choices on their own—due for example to informational or self-control issues.

A small literature in optimal taxation has examined the desirability of education subsidies in fully rational models. Higher education subsidies encourage skill acquisition but tend to benefit more the relatively skilled and hence are likely regressive. Absent any ability to observe educational choices, the total elasticity of earnings with respect to net-of-tax rates is due to both labor supply and education choices. If education choices are elastic, the corresponding optimal income tax should incorporate the full elasticity and not solely the labor supply elasticity. This naturally leads to lower optimal tax rates than those calibrated using solely the labor supply elasticity. Diamond and Mirrlees (Unpublished) develop this point, which they call the "Le Chatelier" principle.[109]

Suppose now that the government can observe educational choices and hence directly subsidize (or tax) them in addition to using income-based taxes and transfers. In that context, redistributive taxes and transfers discourage both labor supply and education investments as they reduce the net rewards from higher education. Bovenberg and Jacobs (2005) consider such a model and show that combining educational subsidies with redistributive income-based taxation is optimal—consistent with real policies.

In the simplest version of their model, education d increases the wage rate $w = n\phi(d)$ (with $\phi(d)$ increasing and concave and n being innate ability) at a cost d. Individuals choose d and l to maximize utility $c - h(l)$ subject to $c = (1 - \tau)n\phi(d) - (1 - s)d + R$ where τ is the income tax rate, s the subsidy rate on education expenses d, and R the demogrant. In this simple model, d is an intermediate good that does not directly enter the utility function which depends solely on c and l. The education choice is given by the first order condition $(1 - \tau)n\phi'(d) = 1 - s$. Hence, education is pure cost of production and individuals should be taxed on their earnings net of education costs $n\phi(d)l - d$. This implies that s should be set exactly equal to τ.

---

[108] The government has better ability than private lenders to enforce repayment of loans based on post-education earnings. For example, in the United States, it is much more difficult to default on (government provided) student loans than on private consumer credit loans.

[109] Related, Best and Kleven (2012) derive optimal tax formulas in a context where effort when young has positive effects on wages later in life.

7. LIMITS OF THE WELFARIST APPROACH AND ALTERNATIVES

7.1. Issues with the Welfarist Approach

All our analysis so far has followed the standard welfarist approach whereby the govern–ment objective is to maximize a weighted sum of individual utilities (or an increasing transformation of utilities). As we saw, all optimal tax formulas can be expressed in terms of the social marginal welfare weights attached to each individual which measure the social value of an extra dollar of consumption to each individual.

In standard optimal tax analysis, the utilitarian case (maximizing the unweighted sum of individual utilities) is by far the most widely used. In that case, social welfare weights are proportional to the marginal utility of consumption. As we have seen, this criterion generates a number of predictions at odds with actual tax systems and with people's intuitive sense of redistributive justice.

First, if individuals do not respond to taxes, i.e., if pre-tax incomes are fixed, and individual utilities are concave, then utilitarianism recommends a 100% tax, and full redistribution. In reality, even absent behavioral responses, many and perhaps even most people would still object to confiscatory taxation on the grounds that people deserve to keep part of the income they have created.

Second and related, views on taxes and redistribution seem largely shaped by views on whether the income generating process is fair and whether individual incomes are deserved or not. The public tends to dislike the redistribution of fairly earned income through one's effort but is in favor of redistributing income earned unfairly or due to pure luck (see Piketty, 1995 for a theoretical model and Alesina & Giuliano, 2011, chap. 4 for a recent survey). Such distinctions are irrelevant for utilitarianism.

Third, as we have seen in Section 6.1 on tagging, under utilitarianism, optimal taxes should depend on all observable characteristics which are correlated with intrinsic earning ability. In practice, taxes and transfers use very few of the potentially available tags. Society seems to have horizontal equity concerns and using tags to achieve indirect redistribution is hence perceived to be unfair.

Fourth, perceptions about recipients seem to matter a great deal for the public views on transfers. Most people support transfers for people really unable to work, such as the truly disabled but most people dislike transfers to people able to work and who would work absent transfers. In the standard model, behavioral responses matter for optimal taxes only through their effects on the government budget. In reality, the presence of behavioral responses also colors the public perceptions on how deserving transfer beneficiaries are.

7.2. Alternatives

A number of alternatives to welfarism have been proposed in the literature.

Pareto Principle. First, let us recall that the standard utilitarian criterion can be easily extended, as we have seen, by considering a weighted sum of individual utilities (instead of a simple sum). Those positive weights are called *Pareto weights*. By changing those

weights, we can describe the set of all second-best Pareto efficient tax equilibria. It seems natural that any "optimal tax system" should be at least second-best Pareto efficient, i.e., no feasible tax reform can improve the welfare of everybody. Hence, the Pareto principle imposes a reasonable but weak condition on tax optima. Indeed, optimal tax analysis was particularly interested in finding properties that hold true for all such second-best optima.[110] Those properties are relatively few, an example being the Atkinson and Stiglitz theorem. Hence, considering arbitrary weights is not going to be enough to obtain definite conclusions in general. Hence, it is necessary to be able to put more structure on those Pareto weights so that we can select among the wide set of second-best Pareto optimal tax systems.

All the examples of alternatives to utilitarianism we describe next show that any criterion leads to a specific set of marginal social welfare weights.

Rawlsian Criterion. In the Rawlsian criterion, Pareto weights are concentrated solely on the most disadvantaged person in the economy. This amounts to maximizing the utility of the person with the minimum utility, hence this criterion is also called the *maxi-min* objective. A judgment needs to be made as to who is the most disadvantaged person. In models with homogeneous preferences and heterogeneous skills, the most disadvantaged person is naturally the person with the lowest skill and hence the lowest earnings. This criterion has the appealing feature that, once society agrees on who is the most disadvantaged person, the optimum is independent of the cardinal choice for individual utilities. The key weakness of this criterion is that it concentrates all social welfare on the most disadvantaged and hence represents extreme redistributive tastes. Intuitively, it seems clear that the political process will put weight on a broader set of voters than solely the most disadvantaged. Hence, the Rawlsian principle makes sense politically only if the most disadvantaged form a majority of the population. This is not a realistic assumption in the case of redistribution of labor income.[111] For example, we have seen in Section 4.1 that a standard median voter outcome puts all the weight on the median voter preferences.

Libertarianism and Benefits Principle. At the other extreme, libertarians argue that the government should not do any redistribution through taxes and transfers. Therefore, taxes should be set according to the benefits received from government spending, individual by individual. This is known as the *benefits principle* of taxation. Any redistribution over and above benefits is seen as unjust confiscation of individual incomes. Such a principle can be formally captured by assuming that social marginal welfare weights are

---

[110] Guesnerie (1995) studies the structure of Pareto optima in the Diamond and Mirrlees (1971) model of linear commodity taxation and Werning (2007) studies the structure of Pareto optima in the Mirrlees (1971) model of nonlinear optimal income taxation.

[111] It is a more realistic assumption in the case of inheritance taxation where indeed about half of the population receives negligible inheritances (see Piketty and Saez (2012a,b) for an analysis of optimal inheritance taxation along those lines).

identical across individuals (in the situation where taxes correspond to benefits). In that case, additional redistribution does not add to social welfare.[112] While some voters may hold libertarian views, as we discussed in Section 2.1, all OECD countries do accomplish very substantial redistribution across individuals, and hence depart very significantly from the benefits principle of taxation. This shows that the benefits principle cannot by itself account for actual tax systems.

Principles of Responsibility and Compensation. The general idea is that individuals should be compensated for circumstances affecting their welfare over which they have no control, such as their family background or disability at birth. This is the principle of compensation. In contrast, individuals should be held responsible for circumstances which they control such as how many hours they work. Hence, no redistribution should take place based on such choices. This is the principle of responsibility. These principles are presented and discussed in detail in Kolm (1996), Roemer (1998), Fleurbaey (2008), Fleurbaey and Maniquet (2011).

An example often presented in the literature is that of individuals differing by their wage rate which they do not control (for example because it is due to exogenous ability), and by their taste for leisure (some people prefer goods consumption, some people prefer leisure consumption). By the principle of compensation, it is fair to redistribute from high wage to low wage individuals. By the principle of responsibility, it is unfair to redistribute from goods lovers toward leisure lovers. When there is only one dimension of heterogeneity, those principles are easy to apply. For example, if individuals differ only according to their wage rate (and not in their tastes), then the principle of compensation boils down to a Rawlsian criterion whereby the tax and transfer system should provide as much compensation as possible to the lowest wage people. In terms of welfarism, social marginal welfare weights are fully concentrated on the lowest wage person. If individuals differ solely in taste for work, the principle of responsibility calls for no redistribution at all because everybody has the same time endowment that they can divide between work and leisure based on their relative tastes for goods consumption vs. leisure consumption. It would be unfair to redistribute based on tastes.[113] The standard welfarist approach cannot easily obtain this meaningful result, except through a renormalization of Pareto weights so that social marginal utilities of consumption are the same across individuals (absent transfers).[114]

---

[112] Weinzierl (2012) proposes a formalization of this principle and considers mixed utilitarian and libertarian objectives. Feldstein (2012) argues that it is "repugnant" to put zero asymptotic welfare weight on top earners (as implied by the utilitarian framework used in the Mirrlees Review), but does not propose an explicit model specifying how the proper welfare weights should be set.

[113] This becomes clear when one considers an equivalent model where everybody has the same money endowment to divide between two goods, say apples and oranges. In such an economy, there is no reason to discriminate in favor of or against apple lovers vs. orange lovers.

[114] Lockwood and Weinzierl (2012) explore the effects of taste heterogeneity for optimal income taxation and show that it can substantially affect optimal tax rates through its effects on social marginal welfare weights.

However, those two principles can conflict in situations where there is heterogeneity in both dimensions (skills and taste for leisure). Fleurbaey (2004) presents a simple example in a two skill, two levels of taste for leisure model showing that it is not possible to fulfill both the responsibility principle and the compensation principle at the same time. Therefore, some trade off needs to be made between the two principles. This trade-off needs to be specified through a social objective function. Fleurbaey (2008) reviews this literature and the many criteria that have been proposed.[115]

Equal Opportunity. One prominent example of how to trade-off the responsibility vs. the compensation principles is Roemer (1998) and Roemer et al. (2003) who propose an *Equal Opportunity* criterion. In the model of Roemer et al. (2003), individuals differ solely in their wage rate w but the wage rate depends in part on family background and in part on merit (i.e., personal effort in getting an education, getting ahead, etc.). The model uses quasi-linear utility functions $u = c - h(l)$ uniform across individuals. In the model, people are responsible for wage differences due to merit but not for wage differences due to family background. Suppose for simplicity there is a low and high family background. The distribution of wage rates is equal to $F_0(w)$ and $F_1(w)$ among those coming from low and high family backgrounds respectively. Assume that high family background provides an advantage so that $F_1(w)$ stochastically dominates $F_0(w)$. The government wants to redistribute from high to low family backgrounds but does not want to redistribute across individuals with different wages within a family background group because their position within the group is due to merit. The government can only observe earnings wl and cannot observe family background (nor the wage rate). Hence, the government is limited to using a nonlinear income tax $T(wl)$ and cannot discriminate directly based on family background. Individuals choose l to maximize their utility $u = wl - T(wl) - h(l)$.

By assumption, two individuals in the same wage percentile p within their family background group are equally deserving. Therefore, any discrepancy in the utility across family background conditional on wage percentile should be corrected. This can be captured by a local social welfare function at percentile p given by $\min_{i=0,1}[w_{p,i}l_{p,i} - T(w_{p,i}l_{p,i}) - h(l_{p,i})]$ where $w_{p,i}$ is the pth percentile wage rate in family background group i, and $l_{p,i}$ the labor supply choice of the pth percentile wage person in group i. Total social welfare is then obtained by summing across all percentiles. Hence, we have

$$SWF = \int_{p=0}^{p=1} \min_{i=0,1}[w_{p,i}l_{p,i} - T(w_{p,i}l_{p,i}) - h(l_{p,i})]dp.$$

Effectively, the social criterion is *locally Rawlsian* as it wants to redistribute across family background groups conditional on merit (percentile) to level the field as much as

---

[115] A number of those criteria can violate the Pareto principle, which is an unappealing feature. Hence, additional axioms have to be added to ensure that the Pareto principle is respected.

possible but does not value redistribution within a family background group (as utilities are quasi–linear).

Because high family background provides an advantage, we have $w_{p,1} > w_{p,0}$. Hence the pth percentile individual in the high family background has a higher utility than the pth percentile individual in the low family background. As a result, total social welfare can be rewritten as:

$$SWF = \int_{p=0}^{p=1} [w_{p,0}l_{p,0} - T(w_{p,0}l_{p,0}) - h(l_{p,0})]dp = \int_{w} [wl - T(wl) - h(l)]dF_0(w),$$

This criterion is equivalent to a standard welfarist objective $\int g(w)[wl - T(wl) - h(l)]dF(w)$ with the following social marginal welfare weights. The weights are equal to zero for those with high family background and equal and constant for those with low family background. Hence, the average social welfare weight at wage w is simply $g(w) = f_0(w)/(f_0(w) + f_1(w))$, i.e., the relative fraction of individuals at wage w coming from a low family background. Presumably, $g(w)$ decreases with w as it is harder to obtain (through merit) a high wage when coming from a low family background.

The standard Diamond (1998) optimal nonlinear tax theory of Section 5 applies in this case by simply substituting the standard welfarist weights by those weights. For example, the optimal top tax rate is given again by the simple formula $\tau = (1 - g)/(1 - g + a \cdot e)$ where g is the relative fraction of top earners coming from a low family background. If nobody coming from a low family background can make it to the top, then $g = 0$ and the optimal top tax rate is set to maximize tax revenue.

Generalized Social Welfare Weights. A systematic approach recently proposed by Saez and Stantcheva (2013) is to consider *generalized social marginal welfare weights* that are ex-ante specified to fit justice principles. Those social marginal welfare weights reflect the relative value of marginal consumption that society places on each individual. Hence, they can be used to evaluate the aggregate social gain or loss created by any revenue neutral tax reform. A tax system is "optimal" if no small revenue neutral reform yields a net gain when adding gains and losses across individuals weighted using those generalized social marginal welfare weights. Importantly, the optimum no longer necessarily maximizes an ex-ante social objective function. Naturally, the optimal tax system that arises is second–best Pareto efficient as long as the social marginal welfare weights are specified to be non–negative.

This framework is therefore general and contains as special cases virtually all the situations we have discussed before. The use of suitable generalized social welfare weights can resolve many of the puzzles of the traditional utilitarian approach and account for existing tax policy debates and structures.

First, if generalized social marginal welfare weights depend positively on net taxes paid, in addition to net disposable income, the optimal tax rate is no longer 100% even absent behavioral responses.

Second, generalized social welfare weights can also capture the fact that society prefers taxes on income due to luck rather than taxes on income due to work. As shown in the example above from Roemer et al. (2003), the social welfare weights can be set to zero for those who have an undue advantage because of family background or income due to luck. Such "locally Rawlsian" weights capture the intuition that it is fair to redistribute along some dimensions but not others. When redistribution is deemed fair, it should be as large as possible as long as it benefits those deem Roemer et al. (2003), Piketty and Saez (2012a,2012b) also use such weights in the context of inheritance taxation where weights are set to zero for all those who receive positive inheritances. In the context of inheritance taxation, this yields relatively robust outcomes, due to the fact that the bottom half of the population generally receives close to zero inheritance. We suspect that this approach could be fruitfully extended to the optimal taxation of top labor incomes. For example, if individuals whose parents were in the bottom half of the income distribution have small probabilities to reach the top 1% of the earnings distribution, then this probability could be used as the welfare weight for the top 1%. One key advantage of this approach–based upon transition probabilities and mobility matrices is that it provides an objective, non–ideological basis upon which welfare evaluations can be made.

Third and related, generalized social welfare weights can capture horizontal equity concerns as well. Weights can be set to zero on anybody who benefits from a favorable treatment based on a policy that creates horizontal inequity (such as, for instance, shorter people in a tax system based on height). In that case, tax policies creating horizontal inequities will arise only if they benefit the group that is being discriminated against, i.e., taxing the tall more is desirable only if the tall end up better off in this new tax system as well. This drastically reduces the scope for using additional characteristics in the tax and transfer system, consistent with the rare use of tags in real policies.

Fourth, generalized social welfare weights can be made dependent on what individuals would have done absent taxes and transfers. For example, social welfare weights can be set to zero on "free loaders" who would have worked absent means–tested transfers. This sharply reduces the desirability of transfers when behavioral responses are large for *fairness reasons* (in addition to the standard budgetary reason).

Naturally, the flexibility of generalized social weights begs the question of what social welfare weights ought to be and how they are formed. First, generalized welfare weights can be derived from social justice principles, leading to a normative theory of taxation. The most famous example is the Rawlsian theory where the generalized social marginal welfare weights are concentrated solely on the most disadvantaged members of society. As we discussed, "locally Rawlsian" weights as in Roemer (1998), Roemer et al. (2003), or Piketty and Saez (2012a,2012b) can also be normatively appealing to model preferences for redistribution based on some but not all characteristics. Second, generalized welfare weights could also be derived empirically, by estimating actual social preferences of the public, leading to a positive theory of taxation. There is indeed a small body of work

trying to uncover perceptions of the public about various tax policies. Those approaches either start from the existing tax and transfers system and reverse engineer it to obtain the underlying social preferences (see e.g., Ahmad & Stern (1984) for commodity taxation and Bourguignon and Spadaro (2012) for nonlinear income taxation) or directly elicit preferences on various social issues in surveys (see e.g., Fong, 2001 and Frohlich & Oppenheimer, 1992). Social preferences of the public are shaped by beliefs about what drives disparities in individual economic outcomes (effort, luck, background, etc.) as in the model of Piketty (1995). In principle, economists can cast light on those mechanisms and hence enlighten public perceptions so as to move the debate back to higher level normative principles.

APPENDIX

A.1 Formal Derivation of the Optimal Nonlinear Tax Rate

We specialize the Mirrlees (1971) model to the case with no income effects, as in Diamond (1998). All individuals have the same quasi–linear utility function $u(c, l) = c - v(l)$ where c is disposable income and l is labor supply with $v(l)$ increasing and convex in l. Individuals differ only in their skill level, denoted by n, which measures their marginal productivity. Earnings are equal to $z = nl$. The population is normalized to one and the distribution of skills is $F(n)$, with density $f(n)$ and support $[0, \infty)$. The government cannot observe skills and thus is restricted to setting taxes as a function only of earnings, $c = z - T(z)$. Individual n chooses l_n to maximize utility $nl - T(nl) - v(l)$ leading to first order condition $n(1 - T'(nl)) = v'(l)$.

Under a linearized income tax system with constant marginal tax rate τ, the labor supply function $l \to l(n(1 - \tau))$ is implicitly defined by the equation $n(1 - \tau) = v'(l)$. Hence $dl/d(n(1 - \tau)) = 1/v''(l)$ and hence the elasticity of labor supply with respect to the net–of–tax rate $1 - \tau$ is $e = (n(1 - \tau)/l)dl/d(n(1 - \tau)) = v'(l)/lv''(l)$. As there are no income effects, this elasticity is both the compensated and the uncompensated elasticity.

Let c_n, $z_n = nl_n$, and u_n denote the consumption, earnings, and utility level of an individual with skill n. The government maximizes a social welfare function,

$$W = \int G(u_n)f(n)dn \quad \text{s.t.} \quad \int c_n f(n)dn \leq \int nl_n f(n)dn - E(p).$$

In the maximization program of the government, u_n is regarded as the state variable, l_n as the control variable, while $c_n = u_n + v(l_n)$ is a function of u_n and l_n. Using the envelope theorem and the individual first order condition, the utility u_n of individual n satisfies $du_n/dn = l_n v'(l_n)/n$.

Hence, the Hamiltonian is

$$H = [G(u_n) + p \cdot (nl_n - u_n - v(l_n))]f(n) + \phi(n) \cdot \frac{l_n v'(l_n)}{n},$$

where $\phi(n)$ is the multiplier of the state variable. The first order condition with respect to l is

$$p\left[n - v'(l_n)\right]f(n) + \frac{\phi(n)}{n} \cdot \left[v'(l_n) + l_n v''(l_n)\right] = 0.$$

The first order condition with respect to u is

$$-\frac{d\phi(n)}{dn} = \left[G'(u_n) - p\right]f(n),$$

which can be integrated to yield $-\phi(n) = \int_n^\infty [p - G'(u_m)]f(m)dm$ where we have used the transversality condition $\phi(\infty) = 0$. The other transversality condition $\phi(0) = 0$ yields $p = \int_0^\infty G'(u_m)f(m)dm$, i.e., social marginal welfare weights $G'(u_m)/p$ average to one.

Using this equation for $\phi(n)$, and noting that $n - v'(l_n) = nT'(z_n)$, and that $[v'(l_n) + l_n v''(l_n)]/n = [v'(l_n)/n][1 + 1/e] = [1 - T'(z_n)][1 + 1/e]$, we can rewrite the first order condition with respect to l_n as:

$$\frac{T'(z_n)}{1 - T'(z_n)} = \left(1 + \frac{1}{e}\right) \cdot \left(\frac{\int_n^\infty (1 - g_m)dF(m)}{nf(n)}\right), \tag{17}$$

where $g_m = G'(u_m)/p$ is the social marginal welfare weight on individual m. This formula is derived in Diamond (1998).

Under a linearized income tax system with marginal tax rate τ, we have $z_n = nl(n(1 - \tau))$ and hence $dz_n/dn = l + (1 - \tau)ndl/d(n(1 - \tau)) = l_n \cdot (1 + e)$. Therefore, denoting by $h(z_n)$ the density of earnings at z_n if the nonlinear tax were replaced by a linearized tax with marginal tax rate $\tau = T'(z_n)$, we have $h(z_n)dz_n = f(n)dn$ and hence $f(n) = h(z_n)l_n(1 + e)$. Therefore, $nf(n) = z_n h(z_n)(1 + e)$ and we can rewrite Eq. (17) as

$$\frac{T'(z_n)}{1 - T'(z_n)} = \frac{1}{e} \cdot \left(\frac{\int_n^\infty (1 - g_m)dF(m)}{z_n h(z_n)}\right) = \frac{1}{e} \cdot \left(\frac{1 - H(z_n)}{z_n h(z_n)}\right) \cdot (1 - G(z_n)), \tag{18}$$

where $G(z_n) = \int_n^\infty g_m dF(m)/(1 - F(n))$ is the average marginal social welfare weight on individuals above z_n. Changing variables from n to z_n, we have $G(z_n) = \int_{z_n}^\infty g_m dH(z_m))/(1 - H(z_n))$ where $H(z_n)$ is the actual (not virtual) cumulative distribution of earnings. This establishes Eq. (11) in the main text. Note that the transversality condition implies that $G(z_0 = 0) = 1$.

Equation (17) is particularly easy to use for numerical simulations calibrated to the actual income distribution. Using the specified utility function $u = c - v(l)$, the distribution $F(n)$ is calibrated so that, using the *actual* tax system, the resulting earnings distribution $H(z)$ match the actual earnings distribution. Once $F(n)$ is obtained, formula (17) can be used iteratively until a fixed point tax system $T'(z_n)$ is found. See e.g., Brewer et al. (2010) for an application to the UK case.

A.2 Optimal Bottom Tax Rate in the Mirrlees Model

In the Mirrlees (1971) model, all individuals have the same utility function $u(c, l)$ increasing in disposable income c and decreasing in labor supply l. Individuals differ only in their skill level, denoted by n, which measures their marginal productivity. Earnings are equal to $z = nl$. The population is normalized to one and the distribution of skills is $F(n)$, with density $f(n)$, and support $[0, \infty)$. The government cannot observe skills and thus is restricted to setting taxes as a function only of earnings, $c = z - T(z)$. Individual n chooses l_n to maximize utility $u(nl - T(nl), l)$ leading to first order condition $n(1 - T'(nl_n))u_c + u_l = 0$. Let $c_n, z_n = nl_n$, and u_n denote the consumption, earnings, and utility level of an individual with skill n. Note that $l_0 = 0$ and $c_0 = -T(0)$.

To have a fraction of non-workers, we assume that $u_l(c, l = 0) > 0$ for all $c \geq 0$. As a result, all individuals with skill n below n_0 defined as $n_0(1 - T'(0))u_c(c_0, 0) + u_l(c_0, 0) = 0$ will not work and choose the corner solution $l_n = 0$ and $c_n = c_0 = -T(0)$. Hence, the fraction non-working in the population is $F(n_0)$ and naturally depends on both $1 - T'(0)$ (substitution effects) and $-T(0)$ (income effects).

Using the envelope theorem, the utility u_n of individual n satisfies $du_n/dn = -l_n u_l/n$. Note that this equation remains true even for non-workers at the bottom as $u_n = u(-c_0, 0)$ is constant with n and hence $du_n/dn = 0$ for $n \leq n_0$.

The government maximizes a social welfare function,

$$W = \int G(u_n) f(n) \, dn \quad \text{s.t.} \quad \int c_n f(n) \, dn \leq \int nl_n f(n) \, dn - E(p).$$

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 $u_n = u(c_n, l_n)$. The Hamiltonian is

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

where $\phi(n)$ is the multiplier of the state variable. As $\partial c/\partial l = -u_l/u_c$, the first order condition with respect to l is

$$p \left(n + \frac{u_l}{u_c} \right) f(n) + \frac{\phi(n)}{n} \cdot \left(-u_l - l_n u_{ll} + l_n u_{cl} \frac{u_l}{u_c} \right) = 0.$$

At $n = n_0$, $l = 0$, $n_0 + u_l/u_c = n_0 T'(0)$, and this first order condition becomes

$$p n_0 f(n_0) T'(0) = \frac{\phi(n_0) u_l}{n_0}.$$

As $\partial c/\partial u = 1/u_c$, the first order condition with respect to u is

$$-\frac{d\phi(n)}{dn} = \left[G'(u_n) - \frac{p}{u_c} \right] f(n) - \phi(n) \frac{l_n u_{cl}}{n u_c}.$$

For $n \leq n_0$, $l_n = 0$, $u_n = u(c_0, 0)$, $u_c = u_c(c_0, 0)$ are constant with n so that this equation simplifies to:

$$-\frac{d\phi(n)}{dn} = \left[G'(u_0) - \frac{p}{u_c}\right]f(n),$$

and can be integrated from $n = 0$ to $n = n_0$ to yield

$$\phi(n_0) = \frac{p}{u_c}\left[1 - \frac{G'(u_0)u_c}{p}\right]F(n_0),$$

where we have used the transversality condition $\phi(0) = 0$. Replacing this expression for $\phi(n_0)$ into the first order condition for l at $n = n_0$ yields

$$n_0 f(n_0) T'(0) = \frac{u_l}{u_c n_0}\left[1 - \frac{G'(u_0)u_c}{p}\right]F(n_0) = (1 - T'(0))\left[\frac{G'(u_0)u_c}{p} - 1\right]F(n_0),$$

which can be rewritten as

$$\frac{T'(0)}{1 - T'(0)} = (g_0 - 1) \cdot \frac{F(n_0)}{n_0 f(n_0)} \quad \text{or} \quad T'(0) = \frac{g_0 - 1}{g_0 - 1 + \frac{n_0 f(n_0)}{F(n_0)}}, \tag{19}$$

where $g_0 = G'(u_0)u_c/p$ is the social marginal welfare weight on non-workers.[116]

Recall that $n_0(1 - T'(0))u_c(c_0, 0) + u_l(c_0, 0) = 0$ which defines $n_0(1 - T'(0), c_0)$. Hence, the substitution effect of $1 - T'(0)$ on n_0 (keeping c_0 constant) is such that $\partial n_0/\partial(1 - T'(0)) = -n_0/(1 - T'(0))$. Hence, the elasticity of the fraction non-working $F(n_0)$ with respect to $1 - T'(0)$ is

$$e_0 \equiv -\frac{1 - T'(0)}{F(n_0)} \frac{dF(n_0)}{d(1 - T'(0))}\bigg|_{c_0} = -\frac{1 - T'(0)}{F(n_0)} \cdot f(n_0) \cdot \frac{\partial n_0}{\partial(1 - T'(0))} = \frac{n_0 f(n_0)}{F(n_0)},$$

which allows us to rewrite (19) as

$$T'(0) = \frac{g_0 - 1}{g_0 - 1 + e_0},$$

exactly as in the discrete model formula (14) presented in the text.

Note that with quasi-linear iso-elastic preferences of the form $u(c, l) = c - l^{1+e}/(1+e)$, the individual first order condition is $[n(1 - T')]^e$ so that everybody with $n > 0$ works. If there is a positive fraction of individuals with zero skill (and hence not working), the formula above applies with $e_0 = 0$ so that $T'(0) = 1$. Intuitively, the fraction of individuals affected by a change in $T'(0)$ is negligible relative to the number of non-workers so that behavioral responses are negligible and hence $e_0 = 0$.

---

[116] Mirrlees (1971), Eq. (44), p. 185 came close to this equation but failed to note the key simplification for one of the terms (ψ_y in Mirrlees' notation) at the bottom when labor supply is zero.

ACKNOWLEDGMENTS

REFERENCES

Adema, W., Fron, P., & Ladaique, M. (2011). Is the European welfare state really more expensive? Indicators on social spending, 1980–2012; and a manual to the OECD social expenditure database. *OECD social, employment and migration working papers*, No. 124.

Ahmad, E., & Stern, N. (1984). The theory of reform and Indian direct taxes. *Journal of Public Economics, 25*, 259–298.

Akerlof, G. (1978). The economics of "tagging" as applied to the optimal income tax, welfare programs, and manpower planning. *American Economic Review, 68*(1), 8–19.

Alesina, A., & Giuliano, P. (2011). Preferences for redistribution. In A. J. Bisin, Benhabib (Ed.). *Handbook of Social Economics* (pp. 93–132). Amsterdam: North Holland.

Alvaredo, F., Atkinson A., Piketty, T., & Saez, E. (2011). The World top incomes database. Online at http://g-mond.parisschoolofeconomics.eu/topincomes/.

Ardant, G. (1971). *Histoire de l'impôt (Vols. 1–2)* (p. 1971). Paris: Fayard.

Atkinson, A. (1995). *Public economics in action*. Oxford: Clarendon Press.

Atkinson, A. & Leigh, A. (2010). Understanding the distribution of top incomes in five Anglo-Saxon countries over the twentieth century. *IZA discussion paper*, No. 4937, May.

Atkinson, A., & Stiglitz, J. E. (1976). The design of tax structure: Direct versus indirect taxation. *Journal of Public Economics, 6*(1–2), 55–75.

Atkinson, A., & Stiglitz, J. E. (1980). *Lectures in public economics*. New York: McGraw Hill.

Atkinson, A., Piketty, T., & Saez, E. (2011). Top incomes in the long-run of history. *Journal of Economic Literature, 49*(1), 3–71.

Auerbach, A. (1988). Capital gains taxation in the United States. *Brookings Papers on Economic Activity, 2*, 595–631.

Auerbach, A., & James, H. (2002). Taxation and economic efficiency. In A. Auerbach, & M. Feldstein (Eds.), *Handbook of public economics (1st ed.). Vol. 3* (pp. 1347–1421). Amsterdam: North-Holland.

Bane, M. J., & Ellwood, D. T. (1994). *Welfare realities: From rhetoric to reform*. Cambridge: Harvard University Press.

Bebchuk, L., & Fried, J. (2004). *Pay without performance: The unfulfilled promise of executive compensation*. Cambridge: Harvard University Press.

Bentham, J. 1791. *Principles of morals and legislation*. London: Doubleday.

Besley, T., & Coate, S. (1992). Workfare versus welfare: Incentives arguments for work requirements in poverty-alleviation programs. *American Economic Review, 82*, 249–261.

Best, M. & Kleven, H. 2012. Optimal income taxation with career effects of work effort. *LSE working paper*.

Blau, F., & Kahn, L. (2007). Changes in the labor supply behavior of married women: 1980–2000. *Journal of Labor Economics, 25*, 393–438.

Blundell, R., & MaCurdy, T. (1999). Labor supply: A review of alternative approaches. In O. Ashenfelter, D. Card (Eds.), *Handbook of labor economics. Vol. 3*. Amsterdam: North-Holland.

Boadway, R. (2012). *From optimal tax theory to tax policy: Retrospective and prospective views*, 2009 Munich lectures in economics. Cambridge, MA: MIT Press.

Boskin, M. J., & Sheshinski, E. (1978). Optimal redistributive taxation when individual welfare depends upon relative income. *Quarterly Journal of Economics, 92*(4), 589–601.

Boskin, M. J., & Sheshinski, E. (1983). Optimal tax treatment of the family: Married couples. *Journal of Public Economics, 20*(3), 281–297.

Bourguignon, F., & Spadaro, A. (2012). Tax-benefit revealed social preferences. *Journal of Economic Inequality, 10*(1), 75–108.

Bovenberg, A. L., & Goulder, L. H. (2002). Environmental taxation and regulation. In A. Auerbach, M. Feldstein (Eds.), *Handbook of Public Economics (1st ed.). Vol. 3* (pp. 1471–1545). Amsterdam: North-Holland.

Brewer, M., Saez, E., & A. Shephard. 2010. Means-testing and tax rates on earnings. In *Dimension of tax design: The mirrlees review, institute for fiscal studies* (pp. 90–173). Oxford University Press.

Cagé, J., & Gadenne, L. (2012). The fiscal cost of trade liberalization. *working paper*, Harvard and PSE.

Card, D., Mas, A., Moretti, E., & Saez, E. (2012). Inequality at work: The effect of peers salary on job satisfaction. *American Economic Review, 102*(6), 2981–3003.

Chetty, R. (2006). A new method of estimating risk aversion. *American Economic Review, 96*(5), 1821–1834.

Chetty, R. (2008). Moral hazard vs. liquidity and optimal unemployment insurance. *Journal of Political Economy, 116*(2), 173–234.

Chetty, R. (2009a). Sufficient statistics for welfare analysis: A bridge between structural and reduced-form methods. *Annual Review of Economics, 1*, 451–488.

Chetty, R. (2009b). Is the taxable income elasticity sufficient to calculate deadweight loss? The implications of evasion and avoidance. *American Economic Journal: Economic Policy, 1*(2), 31–52.

Chetty, R. (2012). Bounds on elasticities with optimization frictions: A synthesis of micro and macro evidence on labor supply. *Econometrica, 80*(3), 969–1018.

Chetty, R., Friedman, J., & Saez, E. forthcoming. Using differences in knowledge across neighborhoods to uncover the impacts of the EITC on earnings. *NBER working paper*, No. 18232. *American Economic Review.*

Chiappori, P. (1988). Rational household labor supply. *Econometrica, 56*(1), 63–90.

Christiansen, V., & Tuomala, M. (2008). On taxing capital income with income shifting. *International Tax and Public Finance, 15*, 527–545.

Congdon, W., Mullainathan, S., & Schwartzstein, J. (2012). A reduced form approach to behavioral public finance. *Annual Review of Economics, 4*, 511–540.

Crépon, B., Duflo, E., Gurgand, M., Rathelot, R. & Zamora, P. (in press). Do labor market policies have displacement effect? Evidence from a clustered randomized experiment. *NBER working paper*, No. 18597. *Quarterly Journal of Economics.*

Cuff, K. (2000). Optimality of workfare with heterogeneous preferences. *Canadian Journal of Economics, 33*(1), 149–174.

Currie, J., & Gahvari, F. (2008). Transfers in cash and in-kind: Theory meets the data. *Journal of Economic Literature, 46*(2), 333–83.

Deaton, A. (1979). Optimally uniform commodity taxes. *Economic Letters, 2*, 357–361.

Delalande, N. (2011a). *Les Batailles de l'impôt. Consentement et résistances de 1789 à nos jours.* Paris Seuil, coll. L'Univers historique.

Delalande, N. (2011b). La Réforme Fiscale et l'Invention des Classes Moyennes–l'Exemple de la Création de l'Impôt sur le Revenu. In P. Bezes, A. Siné (Eds.), *Gouverner (par) les Finances Publiques.* Paris: Presses de Sciences Po.

Diamond, P. (1975). A many-person ramsey tax rule. *Journal of Public Economics, 4*(4), 335–342.

Diamond, P. (1980). Income taxation with fixed hours of work. *Journal of Public Economics, 13*, 101–110.

Diamond, P. (1998). Optimal income taxation: An example with a U-shaped pattern of optimal marginal tax rates. *American Economic Review, 88*, 83–95.

Diamond, P., & Mirrlees, J. (1971). Optimal taxation and public production I: Production efficiency and II: Tax rules. *American Economic Review, 61* 8–27 and 261–278.

Diamond, P. & Mirrlees, J. (Unpublished). *Optimal Taxation and the Le Chatelier Principle.* MIT working paper.

Diamond, P., & Saez, E. (2011). The case for a progressive tax: From basic research to policy recommendations. *Journal of Economic Perspectives, 25*(4), 165–190.

Diamond, P., & Sheshinski, E. (1995). Economic aspects of optimal disability benefits. *Journal of Public Economics, 57*, 1–23.

Duflo, E. (2003). Grandmothers and granddaughters: Old-age pensions and intra-household allocation in South Africa. *World Bank Economic Review, 17*, 1–25.

Dupuit, J. (1844). On the measurement of the utility of public works translated. In K. J. Arrow, T. Scitovsky (Eds.), *Readings in welfare economics* (1969), London: Allen and Unwin.

Eaton, J., & Rosen, H. S. (1980). Optimal redistributive taxation and uncertainty. *Quarterly Journal of Economics, 95*, 357–364.

Edgeworth, F. Y. (1897). The pure theory of taxation. *Economic Journal, 7*, 46–70, 226–238, and 550–571.

Eurostat (2012). *Taxation trends in the European union*. Luxembourg: Publications Office of the European Union.

Fehr, E., & Schmidt, K. M. (1999). A theory of fairness, competition, and cooperation. *Quarterly Journal of Economics, 114*(3), 817–868.

Feldstein, M. (1995). The effect of marginal tax rates on taxable income: A panel study of the 1986 tax reform act. *Journal of Political Economy, 103*(3), 551–572.

Feldstein, M. (1999). Tax avoidance and the deadweight loss of the income tax. *Review of Economics and Statistics, 81*(4), 674–680.

Feldstein, M. (2012). The mirrlees review. *Journal of Economic Literature, 50*(3), 781–790.

Fisher, I. (1919). Economists in public service: Annual address of the president. *American Economic Review, 9*(1), 5–21.

Fleurbaey, M. (2004). On fair compensation. *Theory and Decision, 36*, 277–307.

Fleurbaey, M. (2008). *Fairness, Responsibility and Welfare*, Oxford: Oxford University Press.

Fleurbaey, M., & Maniquet, F. (2011). *A theory of fairness and social welfare*. Cambridge: Cambridge University Press.

Flora, P. (1983). *State, economy, and society in western Europe. Vol. I.* (pp. 1815–1975). London: Macmillan Press.

Fong, C. (2001). Social preferences, self-interest, and the demand for redistribution. *Journal of Public Economics, 82*(2), 225–246.

Frohlich, N., & Oppenheimer, J. A. 1992. *Choosing justice: An experimental approach to ethical theory*. Berkeley University of California Press, Berkeley: Berkeley University Press.

Gauthier, A. H. (2007). The impact of family policies on fertility in industrialized countries: A review of the literature. *Population Research and Policy Review, 26*(3), 323–346.

Golosov, Michael, Tsyvinski, Aleh, & Ivan Werning. 2006. New dynamic public finance: A user's guide. NBER Macroeconomics Annual. volume 21, chapter 5, 317–385, Cambridge, MA: MIT Press.

Goolsbee, A. (2000). What happens when you tax the rich? Evidence from executive compensation. *Journal of Political Economy, 108*(2), 352–378.

Gordon, R., & Slemrod, J. (2000). Are real responses to taxes simply income shifting between corporate and personal tax bases? In J. Slemrod (Ed.), *Does atlas shrug? The economic consequences of taxing the rich* (pp. 240–288). New York: Russell Sage Foundation and Harvard University Press.

Guesnerie, R. (1995). *A contribution to the pure theory of taxation*. Cambridge, MA: Cambridge University Press.

Hungerbuhler, M., Lehmann, E., Parmentier, A., Der Linden, V., & Bruno. (2006). Optimal redistributive taxation in a search equilibrium model. *Review of Economic Studies, 73*, 743–767.

Kaplow, L. (2006). On the undesirability of commodity taxation even when income taxation is not optimal. *Journal of Public Economics, 90*(6–7), 1235–50.

Kaplow, L. (2008). *The theory of taxation and public economics*. Princeton: Princeton University Press.

Katz, M. B. (1996). *In the shadow of the poorhouse: A social history of welfare in the United States. 2nd ed..* Basic Books: New York, NY.

Kirchgassner, G., & Pommerehne, W. (1996). Tax harmonization and tax competition in the European union: Lessons from Switzerland. *Journal of Public Economics, 60*, 351–371.

Kleven, H., & Kopczuk, W. (2011). Transfer program complexity and the take up of social benefits. *American Economic Journal: Economic Policy, 3*, 54–90.

Kleven, H., & Schultz, E. (2012). Estimating taxable income responses using danish tax reforms. *LSE working paper*.

Kleven, H., Kreiner, C., & Saez, E. (2009a). The optimal income taxation of couples. *Econometrica, 77*(2), 537–560.

Kleven, H., Kreiner, C., & Saez, E. (2009b). Why can modern governments tax so much? An agency model of firms as fiscal intermediaries. *NBER working paper*, No. 15218.

Kleven, H., Landais, C., Saez, E., & Schultz, E. (2013). Migration and wage effects of taxing top earners: Evidence from the foreigners' tax scheme in Denmark. *NBER working paper*, No. 18885.

Kleven, H., Landais, C., & Saez, E. (in press). Taxation and international mobility of superstars: Evidence from the European football market. *American Economic Review*.

Kocherlakota, N. R. (2010). *The new dynamic public finance*. Princeton: Princeton University Press.

Kolm, S.-C. (1996). *Modern theories of justice*. Cambridge: MIT Press.

Kopczuk, W. (2005). Tax bases, Tax rates and the elasticity of reported income. *Journal of Public Economics, 89*(11–12), 2093–2119.

Landais, C., Michaillat, P., & Saez, E. (2010). Optimal unemployment insurance over the business cycle. *NBER working paper*, No. 16526.

Landais, C., Piketty, T., & Saez, E. (2011). *Pour une révolution fiscale: Un impôt sur le revenu pour le XXIème siècle*. Paris: Le Seuil.

Laroque, G. R. (2005). Indirect taxation is superfluous under separability and taste homogeneity: A simple proof. *Economics Letters, 87*(1), 141–144.

Lee, D., & Saez, E. (2012). Optimal minimum wage in competitive labor markets. *Journal of Public Economics, 96*(9–10), 739–749.

Lindert, P. (2004). *Growing public: Social spending and economic growth since the eighteenth century*. [Two volumes]. Cambridge, MA: Cambridge University Press.

Lochner, L., & Monge-Naranjo, A. (2011). The nature of credit constraints and human capital. *American Economic Review, 101*(6), 2487–2529.

Lockwood, B. B., & Weinzierl, M. C. 2012. De Gustibus non est taxandum: Theory and evidence on preference heterogeneity and redistribution. *NBER working paper*, No. 17784.

Lundberg, S., Pollak, R., & Wales, T. (1997). Do husbands and wives pool their resources? Evidence from the United Kingdom child benefit. *Journal of Human Resources, 32*, 463–480.

Luttmer, E. (2005). Neighbors as negatives: Relative earnings and well-being. *Quarterly Journal of Economics, 120*(3), 963–1002.

Mankiw, N. G., & Weinzierl, M. (2010). The optimal taxation of height: A case study of utilitarian income redistribution. *American Economic Journal: Economic Policy, 2*(1), 155–176.

Mehrotra, A. K. (2005). Edwin R. A. Seligman and the beginnings of the US income tax. *Tax Notes* (pp. 933–950) [November 14].

Mirrlees, J. A. (1971). An exploration in the theory of optimal income taxation. *Review of Economic Studies, 38*, 175–208.

Mirrlees, J. A. (1976). Optimal tax theory: A synthesis. *Journal of Public Economics, 6*, 327–358.

Mirrlees, J. A. (1982). Migration and optimal income taxes. *Journal of Public Economics, 18*, 319–41.

Mirrlees, J. A. (1986). The theory of optimal taxation. In K. J. Arrow, M. D. Intriligator (Eds.), *Handbook of mathematical economics. Vol. 3* (pp. 1197–1249). Amsterdam: North-Holland.

Mirrlees, J. A. (Ed.). (2010). *Dimension of tax design: The mirrlees review*. Institute for Fiscal Studies. Oxford: Oxford University Press.

Mirrlees, J. A. (Ed.). (2011). *Tax by design: The mirrlees review*. Institute for Fiscal Studies, Oxford: Oxford University Press.

Moffitt, R., & Wilhelm, M. (2000). Taxation and the labor supply decisions of the affluent. In J. Slemrod (Ed.), *Does atlas shrug? The economic consequences of taxing the rich* (pp. 193–234). New York: Russell Sage Foundation and Harvard University Press.

Musgrave, R. (1985). A brief history of fiscal doctrine. In A. J. Auerbach, M. Feldstein (Eds.), *Handbook of Public Economics. Vol. 1* (pp. 1–59). Amsterdam: North-Holland.

Naito, H. (1999). Re-examination of uniform commodity taxes under a non-linear income tax system and its implication for production efficiency. *Journal of Public Economics, 71*, 165–188.

Nichols, A., & Zeckhauser, R. (1982). Targeting transfers through restrictions on recipients. *American Economic Review, 72*(2), 372–377.

OECD. (1986). *Personal income tax systems*. Paris: OECD.

OECD. (2005). Increasing financial incentives to work: The role of in-work benefits. In *OECD employment outlook*, OECD, Paris [2005 Edition].

OECD. (2006). Policies targeted at specific workforce groups or labour market segments. In *OECD employment outlook: Boosting jobs and incomes*, OECD, Paris [2006 Edition].

OECD. (2011a). *Revenue statistics*, 1965–2010. OECD, Paris [2011 Edition].

OECD. (2011b). The taxation of low-income workers. In *OECD tax policy study No. 21: Taxation and employment*, OECD, Paris.

OECD. (2011c). The taxation of mobile high-skilled workers. In *OECD Tax Policy Study No. 21: Taxation and employment*, OECD, Paris.

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

Pareto, V. 1896. La courbe de la répartition de la richesse. *Ecrits sur la courbe de la répartition de la richesse* (pp. 1–15) [Writings by Pareto collected by G. Busino, Librairie Droz, 1965].

Persson, T., & Tabellini, G. (2002). Political economics and public finance. In A. J. Auerbach, M. Feldstein (Eds.), *Handbook of public economics. Vol. 3* (pp. 991–1042). Amsterdam: North-Holland.

Piketty, T. (1995). Social mobility and redistributive politics. *Quarterly Journal of Economics, 110*(3), 551–584.

Piketty, T. (1997). La redistribution fiscale face au Chômage. *Revue Française d'Economie, 12*, 157–201.

Piketty, T. (2001). Les hauts revenus en France au 20e siècle—Inégalités et redistributions 1901–1998 (p. 807). Paris: Grasset.

Piketty, T., & Nancy Q. (2009). Income inequality and progressive income taxation in China and India: 1986–2015. *American Economic Journal Applied Economics 1*(2), 53–63.

Piketty, T., & Saez, E. (2003). Income inequality in the United States, 1913–1998. *Quarterly Journal of Economics, 118*(1), 1–39.

Piketty, T., & Saez, E. (2007). How progressive is the US federal tax system? A historical and international perspective. *Journal of Economic Perspectives, 21*(1), 3–24.

Piketty, T., & Saez, E. (2012a). A theory of optimal capital taxation. *NBER Working Paper*, No. 17989.

Piketty, T., & Saez, E. (forthcoming). A theory of optimal inheritance taxation. *CEPR discussion paper*, No. 9241. *Econometrica*.

Piketty, T., Saez, E., & Stantcheva, S. (forthcoming). Optimal taxation of top labor incomes: A tale of three elasticities. *NBER working paper*. No. 17616. *American Economic Journal: Economic Policy*.

Ramey, V. A., & Francis, N. (2009). A century of work and leisure. *American Economic Journal: Macroeconomics, 1*(2), 189–224.

Ramsey, F. (1927). A contribution to the theory of taxation. *Economic Journal, 37*(145), 47–61.

Roemer, J. (1998). *Equality of opportunity*. Cambridge: Harvard University Press.

Roemer, J., et al. (2003). To what extent do fiscal systems equalize opportunities for income acquisition among citizens? *Journal of Public Economics, 87*, 539–565.

Roine, J., Vlachos, J., & Waldenstrom, D. (2009). The long-run determinants of inequality: what can we learn from top income data? *Journal of Public Economics, 93*(7–8), 974–988.

Rothschild, C., & Scheuer, F. (2011). Optimal taxation with rent-seeking. *NBER working paper*, No. 17035.

Sadka, E. (1976). On income distribution, incentive effects and optimal income taxation. *Review of Economic Studies, 43*(1), 261–268.

Saez, E. 1999. *A characterization of the income tax schedule minimizing deadweight burden*. MIT PhD thesis.

Saez, E. (2001). Using elasticities to derive optimal income tax rates. *Review of Economic Studies, 68*, 205–229.

Saez, E. (2002a). Optimal income transfer programs: Intensive versus extensive labour supply responses. *Quarterly Journal of Economics, 117*(2), 1039–73.

Saez, E. (2002b). The desirability of commodity taxation under non-linear income taxation and heterogeneous tastes. *Journal of Public Economics, 83*(2), 217–230.

Saez, E. (2004a). The optimal treatment of tax expenditures. *Journal of Public Economics, 88*(12), 2657–2684.

Saez, E. (2004b). Direct or indirect tax instruments for redistribution: Short-run versus long-run. *Journal of Public Economics, 88*(3–4), 503–518.

Saez, E. (2004c). Reported incomes and marginal tax rates, 1960–2000: Evidence and policy implications. In J. Poterba (Ed.), *Tax policy and the economy. Vol. 18* (pp. 117–174).

Saez, E. & Stantcheva, S. (2013). Generalized social marginal welfare weights for optimal tax theory. *NBER working paper*, No. 18835.

Saez, E., Slemrod, J., & Giertz, S. (2012). The elasticity of taxable income with respect to marginal tax rates: A critical review. *Journal of Economic Literature, 50*(1), 3–50.

Seade, J. K. (1977). On the shape of optimal tax schedules. *Journal of Public Economics, 7*(1), 203–236.

Seade, J. K. (1982). On the sign of the optimum marginal income tax. *Review of Economic Studies, 49*, 637–643.

Seligman, E. R. A. (1911). *The income tax: A study of the history, theory and practice of income taxation at home and abroad*. New York: Macmillan.

Sheshinski, E. (1972). The optimal linear income tax. *Review of Economic Studies, 39*(3), 297–302.

Simula, L., & Trannoy, A. (2010). Optimal income tax under the threat of migration by top-income earners. *Journal of Public Economics, 94*, 163–173.

Slemrod, J. (1996). High income families and the tax changes of the 1980s: The anatomy of behavioral response. In M. Feldstein, J. Poterba (Eds.), *Empirical foundations of household taxation* (pp. 169–192). Chicago: University of Chicago Press.

Slemrod, J., & Kopczuk, W. (2002). The optimal elasticity of taxable income. *Journal of Public Economics, 84*(1), 91–112.

Slemrod, J., & Yitzhaki, S. (2002). Tax avoidance, evasion and administration. In A. Auerbach, M. Feldstein (Eds.), *Handbook of public economics (1st ed.). Vol. 3* (pp. 1423–1470). Amsterdam: North-Holland.

Sorensen, P. B. (1999). Optimal tax progressivity in imperfect labour markets. *Labour Economics, 6*, 435–452.

Stantcheva, S. (2011). Optimal taxation with adverse selection in the labor market. *MIT working paper*.

Stiglitz, J. (1982). Self-selection and Pareto efficient taxation. *Journal of public economics, 17*, 213–240.

Stiglitz, J. (1987). Pareto efficient and optimal taxation and the new new welfare economics. In A. J. Auerbach, & M. Feldstein (Eds.), *Handbook of Public Economics. Vol. 2*, (pp. 991–1042). Amsterdam: North-Holland.

Treasury, US. (2005). *Simple, fair, and pro-growth: Proposals to fix America's tax system*. President's Advisory Panel on Federal Tax Reform: Washington, DC.

Tuomala, M. (1990). *Optimal income tax and redistribution*. Oxford: Clarendon Press.

US Treasury Department, Internal Revenue Service. (2012). Statistics of income: Individual statistical tables by tax rate and income percentile. Table 1 available online at http://www.irs.gov/taxstats/indtaxstats/article/0,id=133521,00.html.

Varian, H. R. (1980). Redistributive taxation as social insurance. *Journal of Public Economics, 14*(1) 49–68.

Vickrey, W. (1945). Measuring marginal utility by reactions to risk. *Econometrica, 13*, 319–333.

Webber, C., & Wildavsky, A. B. (1986). *A history of taxation and expenditure in the western world*. New York: Simon and Schuster.

Weinzierl, M. C. (2011). The surprising power of age-dependent taxes. *Review of Economic Studies, 78*(4), 1490–1518.

Weinzierl, M. C. (2012). Why do we redistribute so much but tag so little? The principle of equal sacrifice and optimal taxation. *Harvard business school working paper*, No: 12-64.

Werning, I. (2007). Pareto efficient income taxation. *MIT working paper*.

Wilson, R. B. (1993). *Nonlinear pricing*. Oxford: Oxford University Press.

Young, C., & Varner, C. (2011). Millionaire migration and state taxation of top incomes: Evidence from a natural experiment. *National Tax Journal, 64*, 255–284.