Charles University in Prague

Faculty of Social Sciences

Institute of Economic Studies

Richard Janoušek

The Principles of Optimal Income Taxation Revisited

Author: **Richard Janoušek**

Supervisor: **PhDr. Martin Gregor PhD.**

Academic Year: **2012/2013**

Declaration of Authorship

I hereby proclaim that I wrote my bachelor thesis on my own under the leadership of my supervisor and that the references include all the resources and literature I have used. I also proclaim that this thesis has not been used to obtain the same or any other degree.

I grant permission to reproduce and to distribute copies of this thesis document in whole or in part for study or research purposes.

Prohlášení

Prohlašuji, že jsem svou práci napsal samostatně a výhradně s použitím citovaných pramenů. Zároveň prohlašuji, že tato práce nebyla využita k získaní jiného nebo stejného titulu.

Souhlasím s tím, aby práce byla zpřístupněna pro studijní a výzkumné účely.

Prague, May 15, 2013

Signature

Acknowledgment

I would like to express my gratitude to PhDr. Martin Gregor, Ph.D., the supervisor of my thesis. It was only due to his support and guidance that I managed to finish this thesis. I would especially like to highlight his willingness to supervise my work despite not being to able to consult with me in person due to my exchange stay.

Poděkování

Na tomto místě bych rád poděkoval PhDr. Martinu Gregorovi, Ph.D., vedoucímu mé práce. Tuto práci jsem byl schopen dokončit jen díky jeho podpoře a vedení. Obzvlástě bych chtěl vyzdvihnout jeho ochotu vést tuto práci, prestože osobní konzultace nebyly možné vzhledem k mému zahraničnímu pobytu.

Bibliografický záznam

JANOUŠEK, Richard. *The Principles of Optimal Income Taxation Revisited* Praha 2013. 60 s. Bakalářská práce, Univerzita Karlova, Fakulta sociálních věd, Institut ekonomických studií. Vedoucí bakalářské práce PhDr. Martin Gregor, Ph.D.

Bibliographic card

JANOUŠEK, Richard. *The Principles of Optimal Income Taxation Revisited* Prague 2013. 60 p. Bachelor thesis, Charles University in Prague, Faculty of Social Sciences, Institute of Economic Studies. Supervisor PhDr. Martin Gregor, Ph.D.

Abstrakt

Tématem této práce je optimální zdanění osob s nejvyššími příjmy. Toto téma již bylo předmětem mnoha dalších akademických prací posledních čtyřiceti let. Doporučení nedanit příjmy těchto osob přitáhlo k tématu pozornost a debata v posledních letech i nadále živě pokračuje. Tento fakt dal také podnět k shrnutí dané problematiky a k diskuzi nejvíce relevantních příspěvků k aspektům zdanění příjmu. Obsahem této práce je nastínění základního informativního pozadí a vysvětlení dvou základních modelů, lineárního a nelineárního. Poté jsou shrnuty dosud nejdůležitější příspěvky, které jsou předmětem diskuze. Na závěr jsou popsány zajímavější a o něco náročnější aspekty. Tato práce poskytuje solidní porozumění tématu a dává podnět pro budoucí práce za účelem nalezení shody.

Abstract

This thesis deals with the topic of optimal income taxation of top earners which has been a topic of numerous academic pieces of work in the past forty years. For instance, a result attracting attention is the zero marginal tax rate for top earners. In the past years, the debate has continued. As a result, the need to summarize the work and discussion on the most relevant aspects of income taxation has emerged.

Therefore in this paper, the essential informational background is given and then the two main models, linear and non-linear, are explained. Then, the debate is summarized and more advanced topics are tackled. Such a paper is needed to provide not only a solid understanding of the topic, but also suggestions for further work in the field to achieve a consensus.

Klíčová slova

optimální zdanění příjmu, tagging, nulová mezní daňová sazba, nejvíce vydělávající osoby, Mirrleesův model

Keywords

optimal income taxation, tagging, zero marginal tax rate, top earners, Mirrlees model

Contents

1 Introduction

The topic of taxation of income is indisputably interesting for numerous reasons, from tax revenue, through transfers, to policy making. Even though significant amount of literature has been devoted to this complex issue, differences in opinions still remain. Therefore, new papers in the field keep being published. Consequently, new approaches are being proposed to deliver a better tax system. The recommendations of the theory often do not resemble practice which could very well be caused by the lack of consensus within the works being published. Not surprisingly, differences also exist in tax systems across individual countries. This is also due to differences among politicians, some of whom advertise low taxes, whereas others higher taxes to pursue equality more intensely. Another reason could be the traditional setup of taxation of the particular country, as some countries have been more redistributive than others. For this and also other reasons, it is useful to see what the theory has to offer.

The aim of this piece of work is then to bring together the relevant literature in order to create a comprehensible summary of the topic of optimal taxation and provide suggestions for further work in the field. More precisely, this paper mainly focuses on top earners, as the significance of their taxation to revenue is understandably higher than of low earners due to the income concentration in the top. One of the newly proposed approaches put under scrutiny is tagging, the idea of making tax rates dependent on individual characteristics.

The paper is structured as follows. First, the essential background is given, from importance of the topic and the time development of tax rates, to the theoretical background needed for the models. Also, the history of theory is summarized in this section. Secondly, the focus is on the basic linear model, its derivation and a comparison of the theory and the actual tax rates. Third, a section on the nonlinear model is provided, focusing on both the top earners and the general tax for a complex picture. Fourth, the main debate and more advanced issues are monitored, including tagging, different forms of tax avoidance, and a normative proposal. Then, the conclusion follows.

2 Essential Background

This section aims to firstly show the importance of the topic and then provide the reader with the basic knowledge on time development in terms of tax rates. This will be done both for some of the main economies globally and for the Czech Republic. Subsequently, the utilitarian approach will be examined, focusing on the social welfare function. Next, the second welfare theorem will be put forward. Then, the topic of labor supply will be discussed, focusing on the intensive margin. Last, but not least, a summary of the theoretical development in the field will be provided.

2.1 Importance of the Topic

There are several ways in which income taxation is of copious importance. Firstly, it is a major contributor to national income, i.e. gross national product without capital depreciation (OECD, 2012c). To be precise, the ratio of tax income to national income has reached about 35% to 50% in a majority of advanced economies in the OECD, with the estimated ratio for the Czech Republic of 35.3% in 2011 (OECD, 2012c). There was a decline during the years 2008 and 2009. However, then the share started to increase again to the current values presented above. To examine the historical development, the pieces of work by Flora (1983) which include data until 1975, combined with the OECD (2012c) report which commences in 1965, serve very well. To shortly summarize the long-term development, at the beginning of the century the ratio averaged at 10%, exhibiting significant increases until the 1970s to about 40%. Since then, the value has not changed considerably. As far as different groups of countries are concerned, advanced economies tend to have the ratio higher than less developed ones.

To comment on taxation of the two main channels of income, labor and capital, differences need to be pointed out. Their contribution to the total revenue from total income taxation varies significantly. The approximate ratio of the revenue from labor to capital income taxation is 3:1, resembling the ratio of labor to capital income (OECD, 2012c). Nonetheless, the ratio changes dramatically when we

approach the top of income distribution. As we approach the top 1 per cent, the share of capital income becomes 54%, further rising to 63% within the top 0.1%, while in the lowest quantile the share is only 9% (Altshuler, Harris, & Toder, 2011). Therefore, it becomes clear that treatment of capital income among top earners is also important to consider. The optimal approach to taxation of different forms of income is discussed further in section 5.4.

When it comes to the spread of top earners, the share of the top percentile exhibited a significant change between the 1970s and the new millennium, more than doubling from less than 10% to about 20% (Piketty & Saez, 2001).[1] However, since then, the top percentile has been hurt by two recessions, the first recession in 2000-2002, the second one 2007-2009, followed by a period of recovery (Saez, 2013). Yet, despite the decrease between 2000 and 2009, we can again witness the share increasing, with a further surge predicted by Saez (2013) due to for instance cuts on the federal tax on large estates. Using the same data as Saez (2013), Figure 1 depicts the time trend of the share of the top decile of the income distribution in the US, as the economic importance of the top decile for potential tax treatment is understandably larger than only of the top percentile.

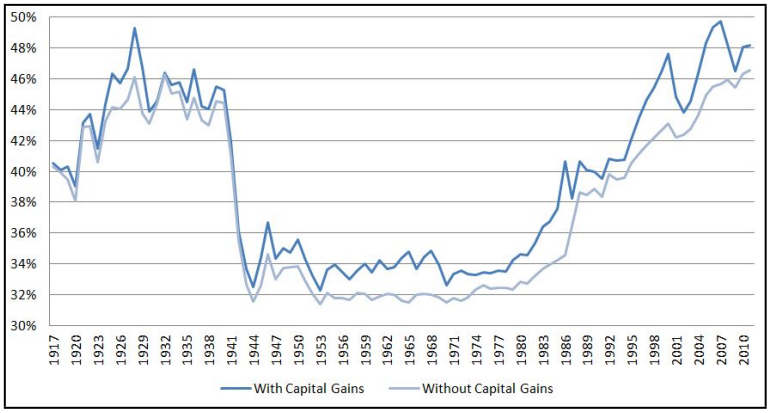

Figure 1: Total Income Share of the Top Decile. Data is shown both with the inclusion and exclusion of capital gains to demonstrate the importance of capital gains and their larger fluctuation, possibly suggesting greater tax elasticity of capital gains (Saez, 2013).[2]

This movement towards income density has been strongest in English speaking

[1] The percentage includes capital gains.

[2] The data can be found in excel format at `http://elsa.berkeley.edu/~saez/TabFig2011prel.xls`.

countries, while continental Europe or Japan have seen only modest shifts (Atkinson, Piketty, & Saez, 2011). As further described in section 2.2, it is useful to bear in mind the simultaneous decrease of top tax rates. The debate on whether there is any causal effect between top tax rates and spread of earners continues to lack a compelling conclusion (see e.g. Piketty, Saez, & Stantcheva, 2011 for a brief overview, or Roine, Vlachos, and Waldenström (2009) for a broader study on income inequality). Nevertheless, from the increased concentration of income, we can see the increased significance of an efficient tax system, especially when it comes to treating top earners. Also notice the grater fluctuation when capital income is included. This shows the importance of bearing all forms of income in mind while creating a tax system, as well as taking into account tax avoidance, an issue further discussed in section 5.3.

Finally, the topic is of utter political importance, for instance British district elections are examined by Revelli (2002). One of the central topics for an election campaign is taxation, especially income taxation. The political note was recently emphasized by France, to be described in section 2.2. The relationship between politics and taxation is empirically explored by Milanovic (2000).

To summarize the prominent aspects of the topic of income taxation, the first to include is the share on national income. The share goes up to 50% in some advanced economies and is above one third in the Czech Republic. Moreover, taxation of top earners is important, as the trend of income concentration is expected to continue. Last, but not least, it is also naturally a topic of interest for voters.

2.2 Development of Tax Systems over the Course of Time

2.2.1 Major and OECD Economies

To begin with, it should be noted that tax systems are often considered jointly with transfers. Many countries have implemented means-tested transfers,[3] so top earners are not their recipients. Consequently, the main focus is on taxes, not

[3]Transfers which are dependent on some test of income. Generally speaking, if income is higher than the level set, then transfers are not given.

transfers.[4] Firstly, we will focus on long-term data on four major economies, namely the United States of America, the United Kingdom, France, and Germany. The beginning of the data is in 1900, we therefore get a solid understanding of the long-term movements affecting the top marginal tax rate. The marginal tax rate is chosen to capture the effect on top earners. Several patterns are observable from Figure 2.

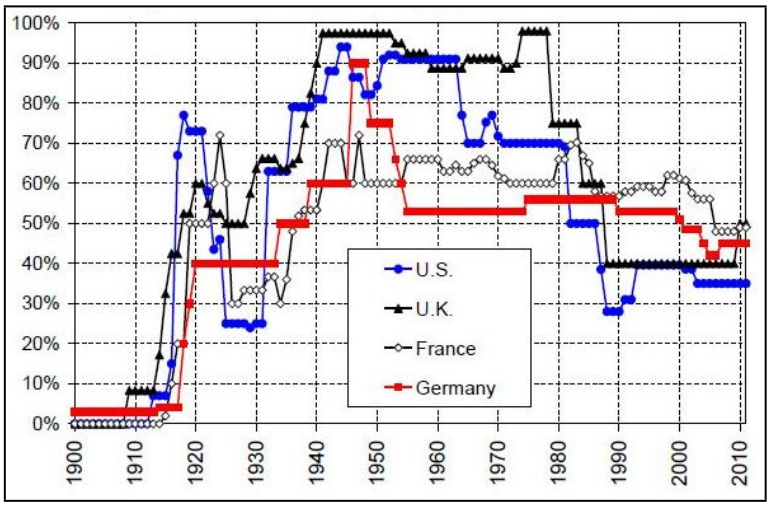

Figure 2: Top Individual Marginal Tax Rates 1900-2011. Taken from Piketty et al. (2011), the tax rate includes only the top statutory individual income tax rate applying to ordinary income with no tax preference. State income taxes are not included in the case of the United States. For France, we include both the progressive individual income tax and the flat rate tax *"Contribution Sociale Généralisée"*.

Firstly, we can notice that upon the implementation of the tax, the marginal rate was relatively low, under 10 per cent. Secondly, the reader should notice the all-time joint peak for all countries in the mid-40s, a result of the rise starting in the mid-1910s. The data then hints that during the World Wars, governments used increases in tax rates to increase their revenue. Since the post-war period, we can see that the top marginal tax rates have been declining. Notice the striking difference in the US top tax rate; in 1950-60s, it exceeded 90 per cent, while today it is 35 per cent.

To see the trend in more countries, we will examine data on OECD countries. The

[4]Yet, there is substantial literature for an interested reader. For a review on the main public spending areas see Adema, Fron, and Ladaique (2011). Trends in transfers are summarized by Cancian and Danziger (2009), and family and marriage taxation by Milligan and Stabile (2008); Kleven, Claus Thustrup Kreiner, and Saez (2006); Alm, Dickert Conlin, and Whittington (1999).

average marginal tax rate on top earners shows a steady diminishing tendency, the average decline between 2000 and 2011 was approximately five per cent (OECD, 2012b; Mankiw & Weinzierl, 2010). To add a recent story on taxation which raised significant controversy, the introduction of the 75 per cent tax on top earners in France serves well as an example. Soon after the implementation, a political storm was caused, some wealthy citizens defected the country and the constitutional court ruled the initial version of the tax bill unconstitutional (Carnegy, 2013). The case still reverberates the news (see e.g. Daneshkhu & Carnegy, 2013), possibly causing the fall in the approval rate of the president Mr. Hollande. This remains the prominent example of recent tax increases in Europe.

Another remarkable development among the OECD countries is the reduction in the number of tax brackets. According to a study of the OECD (2012d), the average number of tax brackets of member states was 14 in 1981. In 1990, the number already dropped to six, further decreasing to five by 2010. This points in the direction of simplification of the previous tax systems. Another movement within the OECD countries is the endorsement of the elimination of international loopholes and standardization of tax systems, as discussed further in section 5.5.

All in all, it can be concluded that the average tax on top earners was relatively low on the verge of the 20$^{\text{th}}$ century and peaked at around the end of World War II. Since then, the average tax rates on top earners have been gradually decreasing, yet remained above the level from the first decades of the previous century. Secondly, the average number of tax brackets among OECD countries has sharply dropped.

2.2.2 Czech Republic

To evaluate the same kind of data for the Czech Republic, the data by the Research Institute for Labour and Social Affairs (2012) is used in Figure 3. From the data, we can clearly see that the trend has resembled the rest of the world by diminishing the top marginal tax rate. The current taxation rate is 15 per cent, making the Czech Republic the country with the second lowest top marginal tax rate in the European Union (Eurostat Press Office, 2012). Nevertheless, Roháč (2013) argues that the tax actually is progressive through social contributions, causing the top

tax rate effectively reaching 22 per cent. The inclusion of social contributions ensures that the Czech taxation system can be more easily compared to other systems, especially when it comes to top income taxation. The future development is of course an entirely different question, conditional on the outcome of the 2014 Parliamentary elections. However, the currently likely-to-win socialist party has implemented into its program an increase of the tax rate to 38 per cent for top earners (Mládek & Havel, 2012).

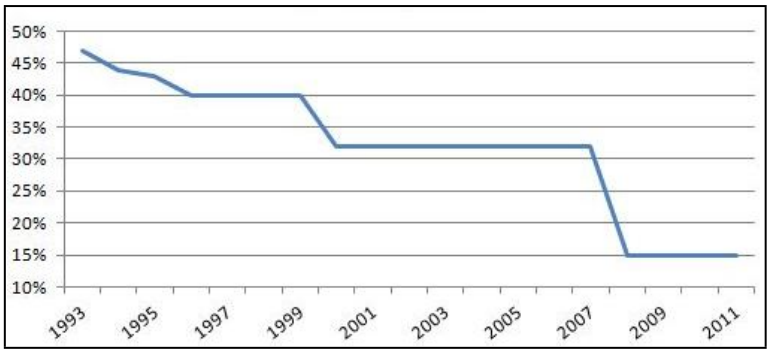

Figure 3: Marginal Tax Rate in the Czech Republic. This figure shows the development of the average marginal top tax rate in the Czech Republic starting with its foundation until today; data by Research Institute for Labour and Social Affairs (2012).

2.3 Social Welfare Function and Second Welfare Theorem

The social welfare function is used very often by theorists to evaluate public policy. One of the simplest methods is to maximize the sum of utilities of individuals within the particular group (nation in case of countries), also called the utilitarian or Benthamite approach (Rosen, 1997). There are two basic approaches to this method, depending on whether earnings are assumed to be fixed or endogenous.

Fixed Earnings This case significantly simplifies the problem due to considerable constraints. To give a brief description following Piketty and Saez (2012), the first assumption is having our population normalized to one. The distribution of pre-tax earnings is also assumed to be fixed, where $H(z)$ is the cumulative distribution function, in other words the fraction of the population earning less than z. Furthermore, the model assumes the utility function $u(c)$ to be homogeneous (identical) across the population, as well as increasing and concave in c which represents disposable income. Time does not play a role here, as one period is

assumed, therefore c is also equal to consumption, such that $c = z - T(z)$ where $T(z)$ is the tax function. The tax function is chosen by the government with the aim to maximize the 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 requirement for the revenue of the government and p is the Lagrange multiplier of the government budget constraint. Bearing in mind that income is fixed, it becomes a point-wise maximization problem.[5] Therefore, the first order condition for $T(z)$ reads:

$$u'(z - T(z)) = p \quad \Rightarrow \quad c = z - T(z).$$

To explain the implication, as p is exogenous, i.e. constant, marginal utility needs to be equalized across z. This is done through equalizing consumption, so $c = z - T(z)$ also has to be constant across z. Explaining the logical reasoning behind equalizing marginal utilities, firstly outlined by Edgeworth (1897), we first have to realize that concave utilities imply decreasing marginal utility, i.e. $u'(c)$ decreases with c. Then $c_1 > c_2$ implies that $u'(c_1) < u'(c_2)$. Therefore, it is desirable to transfer funds from consumer c_1 to c_2. This is achieved by the government imposing a 100% tax on earnings, funding itself and then redistributing the rest of the funds equally between the citizens.

Another version of the social welfare function uses a transformation $G(.)$ which is also increasing and concave in earnings. The generalized functional form is then $\int G(u(c))dH(z)$. If $G(.)$ becomes infinitely concave, we arrive at the *Rawlsian* or *maxi-min* approach. The idea behind this approach is maximizing the minimal utility, thus again redistributing all earnings, exactly as above.

Allowing for heterogeneity of utility functions, we denote $u_i(c)$ as the utility of individual i. While trying to achieve the optimum of identical values of $u_i'(c)$, we face trouble as the comparison marginal utilities becomes challenging due to for

[5] Point-wise means that we examine every value $u(x)$, or every point of some function u.

instance medical expenses, number of dependent children, or simply differences in the joy of consumption (high or low $u_i'(c)$). As in many cases, economic theory starts to struggle here, as we cannot observe properties of individual utility functions. Another aspect is how the society would react to such taxation. All in all, it might be optimal of the government to assume equal marginal utility of consumption across individuals conditional on disposable income. This results in the ability to assume a single general utility function $u(c)$. Its concavity represents the redistributional tastes of the society better than individual marginal utility of consumption.

To summarize, in the case of fixed earnings and homogeneous preferences, maximizing the social welfare function is achieved via full taxation and subsequent redistribution. This ensures that consumption is equal among the population. With heterogeneous utility functions, the problem becomes more challenging due to for instance different joy from consumption. As a result, it might be best to assume equal marginal utility of consumption conditional on disposable income across the population. Then, the preferences can be expressed by a single utility function $u(c)$.

Endogenous Earnings While the result derived in the previous paragraph benefits from simple intuition, the constraint in the form of fixed earnings is rather strong. There are several reasons to relax it. Most importantly, if consumption and hence utility were constant across z, then a rationally behaving agent with pleasure from leisure and disutility from work would have no reason to work. Therefore, the overall level of pre-tax earnings would be at risk. As a result, the general optimization case without such a strong constraint will be considered now. A brief explanation will be given just as for the case of fixed earnings, following Piketty and Saez (2012) who build upon Vickrey (1945); Mirrlees (1971). This is the point where the *equity-efficiency* trade-off comes into play. As the name suggests, on the one hand there is equity across the population, while on the other the (dis)incentives in the form of labor taxation. According to utilitarianism, complete redistribution is then prevented by behavioral responses only. Nevertheless, other

reasons, such as the notion of fairness, may exist in the society. Such issues are further discussed in section 5.8. The bottomline is that the need to consider the case of endogenous earnings is high. The derivations in this section will therefore be of utter importance for the models delivering the tax formulas.

In this section, earnings are a function of labor supply as well, resulting in a slightly different individual utility function $u_i(c, z)$ where z indirectly represents labor supply. The function is increasing in consumption c, but decreasing in earnings z, as higher earnings require higher labor supply. Therefore the above derived 100% taxation result would result in the extinction of work, due to the disutility of labor supply.[6] Last, but not least, the population is again normalized to one which will persist for both the linear and nonlinear model. The social welfare function in this case becomes:

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

where $\omega_i \geq 0$ are *Pareto weights*[7] independent of individuals choices of c and z; $G(.)$ is an increasing transformation of utilities, and $\upsilon(i)$ the distribution of individuals. Due to the usage of arbitrary Pareto weights ω_i and a transformed social welfare function $G(.)$, we can be fully general at this point.[8] The *social marginal welfare weight* on individual i is given by the formula:

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

where p is once again the multiplier of the government budget constraint and u_c^i is marginal utility of consumption. As expected, g_i expresses the value (in terms of public funds) of providing the individual i with 1 additional unit of the

[6]Still, we can consider the case of a very small economy where tax revenue is used for public goods. Then such a 100% tax might not completely extinguish labor. Such an example might be kibbutz communities. However, even they have implemented reforms bringing the tax rates closer to the Israeli ones (Abramitzky & Lavy, 2011).

[7]The Pareto weights used when extending the standard utilitarian criterion through a weighted sum of individual utilities. These weights are positive and via changing them, all second-best Pareto efficient tax equilibria can be described.

[8]Notice that the Pareto weights are not a part of the transformed social welfare function, but only multiply it. This allows for more complex operations. Yet, we can still use this form to evaluate the Rawlsian criterion, as will be shown later.

respective currency. If earnings are fixed and efficiency concerns aside, all g_i's are equalized in the optimum, as again redistribution from those with higher g_i to those with a lower g_i would be desirable to maximize the social welfare function. Following utilitarianism with concave and uniform utility functions, we again arrive at equalization of post-tax income. However, in our case of endogenous earnings, the social marginal welfare weights g_i's ought not to be equalized, as responses of the agents become significantly more important and need to be considered. Societal preferences for redistribution appear in the form of g_i weights in optimal tax formulas.

If we firstly consider the Rawlsian criterion, then it is so that g_i is zero for everyone except for the most disadvantaged person. Taking the utilitarian approach, we will have $g_i = u_c^i/p$ directly proportional to the marginal utility of consumption. Without any income effects on labor supply, the average of g_i across individuals is equal to one and the labor supply decision does not depend on non-labor income.

Before moving on to the next section to explain labor supply, it should be pointed out that while the case of fixed earnings is useful to take into consideration, the main focus is on endogenous earnings as this case with less constraints better represents the reality. Now, another welfare relevant theoretical issue is to be described. It is the Second Welfare Theorem which helps us determine the policy instruments to use in our system.

Second Welfare Theorem The second welfare theorem states that with the standard perfect market assumptions, every Pareto efficient outcome can be reached through a suitable set of lumpsum taxes dependent on exogenous characteristics of individuals (for instance intrinsic abilities, other endowments, or random shocks) and the subsequent free functioning of markets, without any government intervention (Piketty & Saez, 2012). The basic idea behind this approach is an imposition of a tax or transfer which is dependent on their earnings ability (given variation in it and equalizing disposable income) and subsequently not taxing any actual earnings at the margin to avoid market distortions. To be concrete, taxes on individual i depicted T_i would be a function of intrinsic characteristics, but not behavior.

This would lead to equalizing g_i's among the population, where labor supply is chosen individually given T_i's. Before proceeding to the next section, it should be noted that, at least theoretically, lumpsum taxes can be used under virtually any assumptions. However, first-best allocations are then not guaranteed (Barr, 1993). Moreover, those with no or low earnings might not be able to pay such a tax.

The standard model claims the government has to base taxes and transfers on actual earnings, thus distorting earnings and creating efficiency costs. However, with the increased ability of the government to observe information on citizens, including abilities in addition to realized earnings, an alternative to lumpsum taxes called tagging has been proposed to solve these distortions. A closer focus on this matter will be presented in the separate section 5.2.

2.4 Labor Supply

Another significant issue concerning the impact of taxation is labor supply, the decision whether to supply labor and if, then how much. To begin with, it should be noted that in our models, the population is standardized to one for easier operations. Mirrlees (1971) presents a utility function in the form of $u^i(c, z/w^i)$, where the utility function differs purely due to heterogeneous wages w^i, also defining labor supply needed to earn z as $l = z/w^i$. On the other hand, Piketty and Saez (2012) prefer using the utility function of the form $u^i(c, z)$, arguing it is more general and allows for heterogeneity in both preferences and ability. One drawback of this form is that the the opportunity to directly compare preferences for consumption and leisure is not present anymore.

Two general concepts in labor supply exist, the first one being the *Intensive margin*, in other words the decision of how much labor to supply. The second concept, the *Extensive margin*, is the mere decision whether to work or not. Let both be explored in the two next paragraphs.

Intensive margin The intensive margin deals with the case where we already know that individual i is working, thus only deciding how much labor to supply. The utility function in this case becomes $u^i((1-\tau)z+R, z)$, where R is a lump-sum

transfer to every citizen, called the *demogrant*. Consumption is equal to the term $(1-\tau)z + R$, i.e. to after-tax earnings plus the demogrant. The individual then chooses z to maximize his or her utility. The first order condition then becomes:

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

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

In the problem of optimal taxation, formulas for optimal taxes are mostly expressed as a function of elasticities.[9] The impact of $1-\tau$ on z^i expresses the *uncompensated elasticity*[10] of earnings with respect to the *net-of-tax* rate $1-\tau$, as the following formula:

$$e_u^i = \frac{1-\tau}{z^i}\frac{\partial z_u^i}{\partial(1-\tau)}.$$

As Ramey and Francis (2009) have researched, labor supply in the richest countries has declined very slightly when taking into account both men and women. On the other hand, wage rates are now five times of what they were a century ago. Assuming constant preferences for consumption and leisure, Ramey and Francis (2009) arrive at the conclusion that the uncompensated elasticity converges to zero. Yet, we cannot conclude that taxes do not affect labor supply due to the large part of taxes serving as a redistributional component of the tax system. Moreover, the assumption that consumption and leisure preferences have not changed also seems rather strong, as the society seems to gradually strive for more leisure. Secondly, the difficulty of labor has likely declined, at least in developed countries. Thirdly, the study of Ramey and Francis (2009) uses data on market labor, not taking into account household labor which has declined.[11] The conclusion of compensated elasticity close to zero then seems questionable. The issue of labor elasticity is cru-

[9]The elasticity expresses how responsive agents are to changes in disposable income, in our case to changes in taxes.

[10]The compensated elasticity, on the other hand, also subtracts the income effect. The uncompensated elasticity operates with the compensated earnings supply function z_c, see Saez, 2001 for detailed derivation and explanation.

[11]For a review on household labor, see Coltrane (2004).

cial for optimal taxation formulas. Different results with two empirically possible values are summarized later in the text in Table 1, after deriving the optimal tax formulas.

To summarize the effect of the marginal tax rate $T'(z)$, it by definition diminishes the worth of making an extra dollar. Moreover, the income tax level $T(z)$ has an increasing effect on labor supply through the *income effect*, while the substitution effect conditional on $T(z)$ reduces labor supply due to the marginal tax rate $T'(z)$. All in all, taxes (where $T'(z) > 0$ and $T(z) > 0$) have an ambiguous effect on labor supply. On the other hand, transfers (where $T'(z) > 0$ and $T(z) < 0$) have a negative effect on labor supply.

Extensive margin To complete the picture to better represent the reality, it should be also reflected that the mere participation in the labor market brings costs. To name some examples of such costs, transportation, kindergarten or job search related costs serve as proper examples. However, as the extensive margin is not related to top earners, but rather low-income, as transfers are mostly means-tested, only two main points should be noted. As Piketty and Saez (2012) summarize, the first point is the recommendation that a system which considers taxes and transfers jointly should be developed. Secondly, phasing out benefits for citizens is another significant aspect to encourage labor force participation. As Kleven, Landais, Saez, and Schultz (2013b) show, another form of the extensive margin can arise when individuals consider moving to received preferential tax treatment, an issue further examined in section 5.5.

2.5 Theoretical Development in the Field

The problem of optimal taxation these days includes two main approaches - linear and nonlinear taxation. Before moving on to the description of the actual models of income taxation, it is useful to describe the historical development in theory for a complex picture. The linear case does not pose that many technical difficulties. Ramsey (1927) was the first to introduce the problem of linear taxation in terms of elasticities. The first modern treatment of the optimal linear income tax is ac-

counted to Sheshinski (1972). It did not take long to see the relationship between labor supply elasticities and optimal linear tax. However, the incentives to develop it did not become compelling enough until the connection with the nonlinear models. This was carried out by for instance Atkinson (1996), Diamond (1998), or Saez (2001). That was a large impulse for the income tax theorists to use the empirical findings on elasticities and behavior responses concerning taxation.

The most extensively used model for modern theory of optimal taxation is the model of Mirrlees (1971). He delivered the first mathematical model which would allow an analysis and even quantification of the problem. A government budget constraint was introduced, as well as unobserved heterogeneity in ability. The government can observe income which is a function of both ability and effort, thus also ability and effort. However, this observation of ability and effort is only indirect. If the planner favors redistribution from high to low earners, taxes can be based on income, but again causing the equity-efficiency trade-off. As Mankiw, Weinzierl, and Yagan (2009) posit, the strength of the Mirrlees approach is the possibility to consider all feasible tax systems. However, this can sometimes turn into a disadvantage, namely a high level of complexity. For also these reasons, several models have built upon this one. The most surprising finding in the field of tax policy was probably the zero marginal tax rate established by Sadka (1976) and J. K. Seade (1977). The verbal argument behind it, as presented by J. K. Seade (1977), assumes a positive marginal tax rate on the single top earner with income y. The marginal tax discourages the top earner from earning more than y, clearly the desire of this individual to further increase his income is not motivated by need, so the effect of the marginal tax rate could be substantial. If the marginal tax rate on earnings above y was zero, then the efficiency costs would decrease and the individual would exert more effort, possibly reaching higher earnings, thus increasing his consumption and utility. At the same time, it is crucial to realize the same tax revenue would be collected, as the income below y would still be taxed as before.

Another extension of the Mirrlees (1971) model was done by Stiglitz (1982), the

major change being the creation of a discrete model with two skills. The result here was identical to the zero marginal top tax rate on the higher skilled. As only two skills exist in this model, the impact of the zero top tax rate seems even larger, thus contributing to the interest in the matter. The lesson from this model is that the system should be set up so that the high skilled do not want to reduce labor supply and appear as the lower skilled. Yet, in this work instead a model first developed by Piketty (1997) is briefly outlined due to its easier meaningful calibration compared to the Stiglitz (1982) model.

Another shift to the development of both the theory and actual tax policy was due to Atkinson and Stiglitz (1976). They showed that differentiated commodity taxation is not useful when nonlinear taxation of income can be introduced, assuming preferences are separable and homogeneous. This provided the incentive to move towards nonlinear taxation and away from the differentiated commodity taxation model by Diamond and Mirrlees (1971). Concerning the taxation of capital income, the result of Atkinson and Stiglitz (1976) also supported taxation of labor income or consumption instead. Recently, a more straightforward proof of the same result was delivered, namely by Kaplow (2004) and Laroque (2005).

Another approach was presented by Diamond (1980), the main difference from Mirrlees (1971) being mainly the focus on the extensive, instead of the intensive, margin. In this case, the optimal marginal tax rate can be below zero. The model was further extended by Saez (2002) to use elasticities to arrive at a negative marginal tax rate at the bottom of an extensive labor supply model.

Lastly, the link between the actual tax policy and the theory will be outlined. A committee chaired by Professor J. E. Meade was asked to deliver recommendations for the United Kingdom in 1978, as described by Harrison (1979). Another example is a compilation of tax theory with the aim of a tax reform again in the United Kingdom, which was delivered by Mirrlees (2010). The actual tax reform was presented by the Institute for Fiscal Studies (Mirrlees & Stuart, 2011).

To comment on the challenges as seen by Piketty and Saez (2012), one of the crucial unanswered questions are behavioral responses to taxation, particularly in

the long-term. While the short- and medium-term responses are summarized by for example Saez, Slemrod, and Giertz (2012), a deeper knowledge on for instance educational responses would be useful. An analysis in this field is currently carried out by (Abramitzky, 2013). In addition, the utilitarian approach most commonly employed in theory is hardly the approach of politicians and the general public. Moreover, many studies use Pareto weights to assess "second-best" equilibria. Yet, these results are heavily dependent on the Pareto weights. A universal solution robust to changes in the Pareto weights is missing.

To summarize, optimal income taxation has been dealt with for over 70 years. Starting with the work of Ramsey (1927) in optimal linear taxation, the most influential present work was delivered by Mirrlees (1971). Attempts to link the theory to practice have been made, with the notable example the Institute for Fiscal Studies in the United Kingdom.

3 Linear Model

General Guidelines No matter which model for optimal income taxation we select, there are still some general and reasonable guidelines we should bear in mind. As Diamond and Saez (2011) state, the first one is that the economic mechanisms are empirically relevant and first order to the matter considered. Secondly, results should be reasonably robust to changes in modeling assumptions, mainly homogeneity of preferences. Third, the policy recommendations need to be feasible. The first and second conditions lead Piketty and Saez (2012) to take the approach of "sufficient statistics" (a survey on this approach is provided by e.g. Chetty, 2008b). On the implementability, structures which are simple and easily explained to the public are preferred. This condition is a matter of discussion, especially when it comes to tagging. The details of this issue should be postponed until the respective sections, so this part will be based on the existing tax and transfer systems, thus satisfying the third condition. Also, let it be reminded that the population remains normalized to one in both the linear and nonlinear models.

3.1 Basic Model

The linear case is closely connected to the nonlinear case. Therefore, it is reasonable to start by presenting the linear model which is, not surprisingly, simpler. Despite the simpler structure, the key trade-off between equity and efficiency is still captured. Due to the simpler structure, the mathematical derivation will be shown in more detail for better understanding. Let the government have an exogenous non-transfer spending E. The tax system ought to be represented by the linear tax rate of τ and an exogenous demogrant R. Following the explanation by Piketty and Saez (2012), aggregate earnings can be denoted by $Z_u(1-\tau, R)$, a function of the tax rate and the demogrant. The budget constraint is represented by $R+E = \tau Z_u(1-\tau, R)$, defining R implicitly as a function of τ only. As a result, aggregate earnings Z_u can be expressed as a function of only $1-\tau$, taking the form of $Z(1-\tau) = Z_u(1-\tau, R(\tau))$. Then the government is trying to determine the

optimal tax rate τ.[12] The aggregate earnings also express the feasibility of a tax system, as they express what the government can generate through it. This is because the reaction of the economy to changes is described, at least in expectation. To comment on the tax revenue, the famous Laffer curve applies.

To describe it, the government has no tax revenue when $\tau = 0$ and $\tau = 1$. In the latter case, the reason is that no-one decides to work. The revenue maximizing tax rate obviously lies between 0 and 1, not necessarily in the middle as could be concluded from the Laffer curve. This can then be used as an obvious assumption. Concerning the demogrant, two scenarios are possible. Either $R \geq 0$ or $R < 0$. If $R \geq 0$, in other words the transfer to individuals is positive, then we set a combined system of τ and the lumpsum tax $-R$. In the second case where $R < 0$, the government essentially implements a lumpsum tax as well (which is now a positive tax). Then, the tax rate decreases compared to the case of $R \geq 0$, as the amount needed to collect to meet the budget constraint combines the individual tax rate and the lumpsum tax revenue. However, it should be noted again that a negative demogrant would be very difficult to justify, as for instance those with no earnings would hardly be able to pay it.

To derive the revenue maximizing taxation rate τ^*, we also firstly 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, recalling Z is a function of $(1-\tau)$. In order to maximize the revenue, we set the tax rate τ^* such that $Z - \tau\frac{dZ}{d(1-\tau)} = 0$, i.e. $\frac{\tau}{1-\tau}e = 1$. Then τ^* can be rewritten as a function of e and vice versa:

$$\textbf{Revenue maximizing linear tax rate: } \frac{\tau^*}{1-\tau^*} = \frac{1}{e} \text{ or } \tau^* = \frac{1}{1+e}. \quad (1)$$

Even though maximizing the revenue surely is an important aspect, maximizing the social welfare function should be of higher importance for policy makers. However, there is one important lesson we can take from here. Let two tax rates τ_1 and τ_2 raise the same revenue. Then using the Laffer curve, we can also assume that $\tau_1 \in (0, \tau^*)$

[12] An alternative approach would be taking the budget constraint E and tax rate τ as exogenous in order to arrive at the demogrant R. This approach might be useful if the government cannot change the tax rate so decides to estimate the optimal demogrant R.

and $\tau_2 \in (\tau^*, 1)$. In that case, τ_1 dominates τ_2, because the tax distortions are lower in the τ_1 case. In the range $(\tau^*, 1)$ also called the *prohibitive range*, a tax-cut would imply a raise in tax revenues, so the economic effect would outweight the arithmetic effect (Laffer, 2004). Laffer (2004) also discusses the successful policy changes after the policy makers realized they might be behind the revenue maximizing rate τ^*, which could be the case in especially social states. To summarize, the optimal tax rate should lie in the interval $(0, \tau^*)$.

Now, we can move on to the derivation of the optimal tax rate maximizing social welfare following Piketty and Saez (2012). The demogrant to every citizen is $\tau Z(1-\tau) - E$. Consequently, disposable income for individual i can be expressed as $c^i = (1-\tau)z^i + \tau Z(1-\tau) - E$. Then τ should maximize the social welfare function in the form:

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

Through the envelope theorem from the choice of z^i in the utility maximization problem for individual i, the government faces the following first order condition:

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

The first term in the brackets $Z - z_i$ is the pure mechanical effect of changing the tax system (in other words the tax and demogrant), behavioral responses aside. When individual income z^i is less than average income Z, then the effect $Z - z^i$ is positive. The aggregate behavioral response is captured in the term $\frac{dZ}{d(1-\tau)}$, reflecting the efficiency cost of a tax increase. It is an efficiency cost due to the first order negative effect of the responses on tax revenue and no first order positive welfare effect on individuals.[13]

Introducing the aggregate elasticity e and normalizing the social welfare weight,

[13]The term first order effect is given by the first derivation, also it may be understood as the linear effect. The second and other effects are defined analogically.

the first order condition can be expressed as:

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

Rearranging this condition, we arrive at the optimal linear income tax formula in the form of:

$$\textbf{Optimal linear tax rate: } \tau = \frac{1-\bar{g}}{1-\bar{g}+e} \text{ with } \bar{g} = \frac{\int g_i z_i d\upsilon(i)}{Z}. \tag{2}$$

The term $\bar{g}$ represents firstly the average normalized social marginal welfare weight weighted by pre-tax incomes z_i, and secondly the ratio of the average income weighted by individual social welfare weights g_i to the actual average income Z. This implies that $\bar{g}$ measures at which income are social welfare weights concentrated. There are five noteworthy items about formula (2).

Firstly, τ, the optimal tax rate, is a decreasing function of the aggregate elasticity e, as e is a mix of substitution and income effects, because an increase in τ implies an increase in the demogrant $R = \tau Z(1-\tau) - E$. Recalling formula (1) for revenue maximization, e becomes a pure uncompensated elasticity, because the tax does not increase revenue at the margin. Similarly, when $\tau = 0$, then it essentially becomes compensated elasticity, as all taxes are returned at no efficiency cost.

Secondly, the other variable in the formula, $\bar{g}$, also causes τ to decrease when $\bar{g}$ increases. $\bar{g}$ denotes the redistributive preferences of the government. The extreme Rawlsian scenario where the person worse off earns $z_i = 0$ means that the goal of the government is to maximize R, in other words the subsequently split tax revenue, through $\tau^* = \frac{1}{1+e}$. In the opposite case of $\bar{g} = 1$ (where the government does not value redistribution at all), $\tau = 0$, provided a lumpsum tax E can raise enough revenue for the government to function.[14]

Third and connected, given the social welfare weights, the higher the inequality before given taxes τ, the lower $\bar{g}$ and thus the higher the optimal tax rate τ.

[14]If it is not sufficient, for instance for low earners (where $z_i < E$), then τ is set as small as possible such that the revenue just covers the spending, i.e. $\tau Z(1-\tau) = E$.

Perfect equality implies $\bar{g} = 1$, hence $\tau = 0$ and a lumpsum tax of $-R = E$ as optimal. On the other hand, in the case of maximal inequality where one person i earns all income (with zero social marginal welfare weight) and therefore all the others have no earnings, then the revenue maximization tax $\tau^* = \frac{1}{1+e}$ is again optimal.

Fourth, all the variables Z, $\bar{g}$, and e are a function of τ, making it an implicit formula as is usual in optimal tax theory. Using the standard utilitarian social welfare criteria with concave utility of consumption, $\bar{g}$ is an increasing function of τ, as higher taxes imply a lower need for redistribution. Elasticity e is often assumed constant for any Z (i.e. any τ), or at least for small changes in τ.

Finally, formula (2) can also evaluate tax reforms. As the formula suggests, if it holds for the current tax rate that $\tau < \frac{1-\bar{g}}{1-\bar{g}+e}$, then an increase in τ has a positive effect on social welfare. Obviously, lowering τ would increase social welfare in the opposite case. This approach gains strength from not requiring how e and $\bar{g}$ change with τ, as it deals only with local changes.

3.2 Generality of the Formula

The generality of the formula is significant.[15] The formula itself is derived mainly from the parameter e. The parameter of aggregate elasticity incorporates the changes in educational and human capital decisions as a response to tax changes, thus achieving long-run applicability (see Best and Kleven (2012) for a survey of long-term effects of work efforts). For a government to set taxes, the only two variables necessary are e and how it sets normalized marginal welfare weights g^i. Moreover, the same formula applies for discrete populations and labor supply responses along the extensive margin. Last, but not least, if earnings are determined not only by ability and effort, but also by sheer luck as modeled by Varian (1980) and Eaton and Rosen (1980), then the same formula still applies provided the social welfare objective is defined over individual expected utilities. The random earnings model generates the same type of optimal tax formula and captures the

[15]Such generality though takes us back to the implicit form. Yet, it can be resolved with assumptions on the utility function.

same equity-efficiency trade-off.

3.3 Connection of the Model with Actual Tax Rates

Having derived the optimal formulas for income taxation, is it possible to explain the actual tax rates used in countries? For this analysis, a table by Piketty and Saez (2012) will be presented in Table 1 and shortly described.

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

Table 1: Optimal Linear Tax Rate Formula $\tau = (1-g)/(1-g+e)$. Panel A takes $\bar{g}$ as given and subsequently computes τ. The current US situation is the base for the $\bar{g}$ in the median voter model, where z_m denotes median income. On the other hand, panel B takes τ as given and finds the corresponding $\bar{g}$. Adapted from Piketty and Saez (2012).

To comment on the empirically realistic elasticity of 0.25, under the Rawlsian criterion the suggested $\tau = 80\%$ is significantly above all actual tax rates, even in countries of the tax to GDP ratio approaching 50%. The utilitarian criterion CRRA (constant relative risk aversion) delivers $\tau = 61\%$, still higher than the actual tax rates. However, with the median voter model of suggested $\tau = 55\%$, the theory seems to meet the average tax rates in high tax countries. The other way around, deriving $\bar{g}$ from actual tax rates, the American example would imply $\bar{g} = 87\%$ which indicates rather low redistributive tastes. The case of the EU implies $\bar{g} = 75\%$ which is close to the median voter model.

In the case of $e = 0.5$, which is considered rather high, the Rawlsian rate would be 67%, still above any actual rate. The utilitarian approach would result in a rate of 48%, meeting the actual rates in high tax countries (very close to the EU average). Thirdly, the median voter optimum is of 38%, approaching the US model. In the opposite direction, tax rate $\tau = 35\%$ would result in $\bar{g}$ of 75%, close to the median

voter model. Then, $\tau = 50\%$ would imply preferences for redistribution of 50%, close only to the utilitarian approach with elasticity of 0.5.

In both models, we can also notice that the optimal rate for the median voter model is below the optimal rates suggested by the other two models. This result is caused by distribution of income, so the formula then delivers lower rates. The explanation could also be that the median voter does not perceive redistribution very favorably. The possible reasons for such perception are further discussed in section 5.8.

All in all, the tax rates in OECD countries seem to correspond to the results from formula (2), provided realistic and moderately high elasticity and modest social preferences for redistribution, especially those of the median voter.

4 Nonlinear Model

While the linear model does provide a useful insight into the problem of optimal taxation, we should bear in mind that governments have the ability to base taxation on the income, applying progressivity. As mentioned above, the Czech Republic, at least officially, does not exercise progressive taxation. However, as many other countries do, it is undoubtedly interesting to discover the optimal approach to such taxation. To do this, we will focus on the rates and not try to model the whole social system or solve the feasibility of demogrants or aggregate revenues, as this is not the main focus of this work.

Before moving on to the description and the model itself, it should be pointed out that the three conditions from the linear model still apply. To recall, they were empirical relevance of economic mechanisms, robustness of results to changes in modeling assumptions, and feasibility of policy recommendations.

Now, the crucial difference between the linear and nonlinear model is that the government can use a nonlinear tax system $T(z)$, a more complex tax structure than a single linear tax τ with a demogrant. The government then still faces the same goal - setting $T(z)$ to maximize the social welfare function in the form of:

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

where individual i chooses z^i to maximize utility $u^i(z^i - T(z^i), z^i)$. In the case of no earnings, individuals receive a transfer of $-T(0)$. The system $T(z)$ then incorporates both taxes and transfers. The section is divided into a standard derivation of the optimal top tax rates, followed by the derivation at any rate for a complex picture. Yet, before moving on to the derivations, a tool for evaluations of tax reforms will be introduced.

Small Tax Reform Effect A key concept in optimization is the impact of a small change. Similarly to basic microeconomic optimization problems, in optimum the effect of a minor change should be equal to zero. The first order maximizing condition in a nonlinear tax system $T(z)$ maximizing $u^i(z - T(z), z)$ is

$u_c^i \cdot (1 - T'(z)) + u_z^i = 0$, where u_c^i and u_z^i represent the marginal utility of consumption and income, respectively. A small reform $dT(z)$ has the following effect on individual utility u^i:

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

where dz is the behavioral response to the tax reform. The second equality yields from the first order condition, having factorized dz. This is an application of the envelope theorem, as utility is a function of z, so a small change dz should not affect individual utility. The essential part to take from this is that the individual behavioral response dz has no first order effect, the first order change in utility is merely the mechanical effect of the tax reform multiplied by marginal utility of consumption (Piketty & Saez, 2012). However, behavioral responses should still be taken into account, bearing in mind they are not first order.

4.1 Optimal Top Tax Rate

Once again, several assumptions exist to deliver the formula under the standard model, following Piketty and Saez (2012). Firstly, let the top tax rate above a fixed income level z^* be constant equal to τ. Secondly, let q represent the fraction of the population earning above z^*. To pursue the derivation of τ^*, a small variation $d\tau$ is considered. Therefore, an individual i in the top bracket will have to pay extra taxes of $[z^i - z^*]d\tau$. Taking into account social welfare weight again, the loss generated by the $d\tau$ reform is $-g^i \cdot [z^i - z^*]d\tau$, where g^i represents the social marginal welfare weight on individual i. Lastly, there is a behavioral response dz^i, further altering the tax by τdz^i. Using the elasticity of reported income z^i with respect to the net-of-tax rate $1-\tau$, we obtain $dz^i = -e^i z^i d\tau/(1-\tau)$. Combining the three effects on individual i, the net effect on individual i is given by the expression:

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

To examine the aggregate impact on the whole society, we use the fraction of the population in the top tax bracket q to obtain:

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

where the newly introduced variable z denotes the average income in the top bracket, g the average social marginal welfare weight weighted by the top income bracket income $z^i - z^*$, and e the average elasticity weighted by income z^i. The last notation in need of introduction is the tail parameter $a = \frac{z}{z-z^*}$, then we can rewrite $dSWF$ as:

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

This and realizing that $dSWF = 0$ at the optimum finally allows us to capture the optimal top tax rate as:

$$\textbf{Optimal top tax rate: } \tau = \frac{1-g}{1-g+a\cdot e}\ . \tag{3}$$

The derivation and formula are very similar to the linear case. One difference is the presence of the tail parameter $a > 1$. When a=1, then we arrive again at the linear top tax rate formula (2).

At this point, several comments should be made on this formula. Firstly, the lower g, the social marginal welfare weight on the top bracket, the higher the optimal tax. If the society did not value marginal welfare of the top bracket at all, the formula would become $\tau = 1/(1+a\cdot e)$, which is the revenue maximizing formula. The most commonly used approach, which builds upon Mirrlees (1971), with marginal utility of consumption approaching zero implies that as $z^* \to \infty$, g converges to zero.[16]

Secondly, the optimal tax rate is a decreasing function of e, as higher elasticity implies higher efficiency costs. While elasticity includes both income and substi-

[16]Yet, another aspect is e.g. fairness of the system, the normative discussion is in section 5.8.

tution effects, the impact of the income effect should be pointed out. For a given compensated elasticity, the income effect increases the optimal top tax rate because raising the tax reduces disponsable income, thus encouraging labor supply.

Thirdly, the optimal tax rate decreases with $a \geq 1$ (recall that $a \in \langle 1, \infty)$ by definition). The parameter a measures the thinness of the top tail of the income distribution. As Piketty and Saez (2012) claim, empirically a is almost constant as z^* varies in the top tail of the earnings distribution. They further claim that the distribution in the top tail is very closely approximated by a Pareto distribution.[17]

Fourthly, a connection with the zero-top tax rate is shown. The optimal tax τ indeed does reach zero when z^* reaches the level of the single highest earner, sending $a \to \infty$ (Sadka, 1976; J. K. Seade, 1977). However, notice once again that this applies to the very top earner. Moreover, assuming that the level of top earnings is unknown beforehand and drawn randomly from a Pareto distribution, and that the budget constraint is satisfied in expectation, formula (3) remains optimal (Diamond & Saez, 2011).

Finally, formula (3) is rather general and applicable to populations with heterogeneous preferences, regardless whether discrete or continuous.

To summarize, the general top tax rate formula gives the optimal rate as a function of the average social marginal welfare weight, average elasticity and the tail parameter depicting the relationship between the average and lowest income of the bracket. The optimal rate is a decreasing function of elasticity and the tail parameter. Conversely, it increases as the social marginal weight on the top bracket rises.

4.2 Optimal Tax System Schedule

4.2.1 Continuous Model of Mirrlees

This part will focus on delivering the formula for the optimal marginal tax rate $T'(z)$ at any income level z in the general nonlinear model. To obtain the formula,

[17]However, disagreement exists whether Pareto or lognormal distribution is closer approximation of reality. For more details, see section 5.1.

a similar method as in the previous section is used. For simplification (as Atkinson (1996) and Diamond (1998) showed), the income effects are assumed not to exist, causing labor supply to depend merely on the net-of-tax rate $1 - T'(z)$.[18] To show the logic of the derivation, Figure 4 is adapted from Piketty and Saez (2012) wherein the basic framework of the proof lies as well.[19]

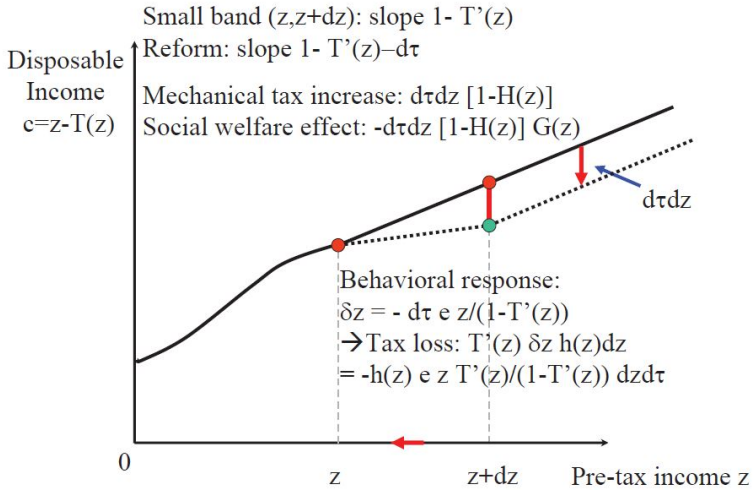

Figure 4: Derivation of the Optimal Marginal Tax Rate at Income Level z (Piketty & Saez, 2012).

As apparent from the figure, pre-tax earnings are depicted on the horizontal axis and disposable income on the vertical axis. We will again consider the case of a small reform in the form of a tax increase $d\tau$ affecting only a small band of people earning between z and $z+dz$. The reform brings three consequences - a mechanical tax increase, a social welfare cost, and a behavioral response.

The mechanical tax increase is $dM = d\tau dz[1 - H(z)]$, because there are $1 - H(z)$ earners above z. In the figure, the pure tax increase (leaving behavioral responses aside) is the gap labeled $d\tau dz$ between the two parallel lines.

The social welfare cost is equal to $dW = -d\tau dz[1 - H(z)]g^+(z)$, as we are only interested in the social marginal welfare weight of individuals above the income z, the function $g^+(z)$ is used to capture its average. Intuitively, dW measures the loss of the society from the decreased welfare resulting from dM.

[18] For the case with income effects, see Saez (2001).

[19] For formal presentation and derivation, see Diamond (1998).

The behavioral response applies only to those in the band $z + dz$ depicted by the horizontal line pointing left. Tax payers in this band reduce their income by $\delta z = -ezd\tau/(1 - T'(z))$ where e is the elasticity of earnings z with respect to the net-of-tax rate $1 - T'$. The total tax loss resulting from the behavioral response is:

$$dB = (h(z)dz) \cdot \delta z \cdot T'(z) = -(h(z)dz)\frac{ezd\tau T'(z)}{1 - T'(z)},$$

as there are $h(z)dz$ taxpayers in the band.[20] Without any income effects, those with income above $z + dz$ have no behavioral response.

If we reach the optimum, the total effect is zero, i.e. $dM + dW + dB = 0$. In addition, we define the local Pareto parameter as $\alpha(z) = zh(z)/(1 - H(z))$. It reflects the ratio of the total income of those affected by the marginal tax rate at z relative to the number of individuals earning more than z. Then, plugging the terms into the equation $dM + dW + dB = 0$, if we simply express $T'(z)$, cancel out $[1 - H(z)]$ and use $\alpha(z)$, we obtain the optimal tax formula in the form of:

$$\textbf{Optimal nonlinear marginal tax rate: } T'(z) = \frac{1 - g^+(z)}{1 - g^+(z) + \alpha(z) \cdot e} \ . \quad (4)$$

As the reader might notice, the formula is strikingly similar to formula (3) for the top tax rate. There are five noteworthy points at this stage (Piketty & Saez, 2012).

Firstly, the graphical proof shows that the formula can be applied also to heterogeneous populations, just as the basic linear tax rate (2).[21]

Secondly, the optimal tax rate is a decreasing function of $g^+(z)$, as the more weight is put on individuals above z, the lower their taxation ought to be. Assuming $\frac{\partial g^+(z)}{\partial z} < 0$, i.e. decreasing social marginal weight with income, and no income effects, the average social marginal welfare weight equals one at $z = 0$. Then $z > 0 \Rightarrow g^+(z) < 1$. As Mirrlees (1971) and J. Seade (1982) showed, then $T'(z) \geq 0$ for any z. The quicker the decreasing pace of $g^+(z)$, the more progressive should

[20] Assuming a locally linear tax schedule. For the nonlinear local case see Saez (2001).

[21] This relaxes the homogeneity assumption of the standard Mirrlees model in which the only difference between individuals is the skills parameter.

the taxes be.[22]

Thirdly, the higher the elasticity e, the larger the efficiency costs in the small band $(z, z + dz)$, thus the lower optimal tax rate.

Fourthly, an increase in the Pareto parameter decreases the marginal tax rate. Empirically in the US, $\alpha(z)$ first rises, then falls and becomes virtually constant for top earners. Therefore, in the top tail it approaches the optimal top rate formula (3).

Fifth, if the government had no redistributive intentions, but only wanted to satisfy its budget needs while keeping efficiency costs to a minimum, formula (4) still applies (with constant $g^+(z)$ set to meet the budget demands) (Saez, 1999).[23]

To summarize the five points made, it is important to realize that assuming constant e, diminishing marginal social weight with income $\frac{\partial g^+(z)}{\partial z} < 0$, and virtually constant $\alpha(z)$ in the top tail, the formula suggests tax progression at the top.

All in all, numerical simulations by Saez (2001) imply that the Pareto parameter effect dominates at the bottom. This implies that the marginal tax rate is high, but decreasing for low earners. Assuming constant elasticity, formula (4) suggests a U-shaped profile of marginal tax rate, first decreasing, then increasing again to the limit given by the optimal tax rate formula (3).

4.2.2 The Discrete Case

Discrete models, which were already briefly described in section 2.5, gained popularity due to their relative simplicity compared to the continuous model. In this part, only the elementary ideas will be outlined. The model of Mirrlees (1971) was extended to deliver a discrete model by Stiglitz (1982). The model incorporates two kinds of workers – high or low skilled, implying high and low wage rates. As Piketty and Saez (2012) posit, it might be deceiving with respect to tax progressivity for two main reasons. Firstly, due to the result of zero top marginal tax rate, it

[22] In the extreme Rawlsian case, $g^+(z) = 0$ everywhere except for $z = 0$. Then $g^+(z)$ drops out and the goal is to maximize tax revenue and thus also the demogrant $-T(0)$.

[23] As those without earnings could hardly pay taxes, this option is not considered.

suggests that taxes should decrease with earnings. Secondly, due to the impossibility to express the Stiglitz (1982) formulas in estimable statistics, it is impossible to calibrate the model. Therefore, Piketty (1997) created and Saez (2002) extended a model with a finite number of earnings levels $z_0 = 0 < z_1 < ... < z_N$. At the same time, there is a continuum of individual types. This ensures that the fraction of individuals at each earning level is a smooth function of the tax system. Then, formulas close to the Mirrlees continuous model are delivered.

Leaving the derivation aside, it should be pointed out that all kinds of behavioral responses are allowed in this model. Two particular kinds of behavioral responses have gained interest: pure intensive responses as in the aforementioned model, and pure extensive responses. To comment on the importance for the whole system, accounting for both these behavioral responses is needed for the optimal determination of transfers.

All in all, the discrete models seem to have undergone development from two skilled populations towards more precise models which better capture the reality. They then deliver results close to the more complex continuous models, while allowing for any behavioral responses.

5 Debated Issues in Income Taxation

Obviously, optimal income taxation of top earners is still a field which has yet to arrive at a consensus. Even though the basic framework provided by Mirrlees (1971) is generally accepted, significantly different recommendations are provided by the theory. The field of optimal taxation appears to have kept its momentum, as new pieces of work are still being published. The focus of this section will lie in mapping the main areas of the debate of economists and approaching more advanced issues. Firstly, the recent discussion between two main groups, represented N. Gregory Mankiw and Peter A. Diamond respectively, will be mapped. Second and related, the notion of tagging is examined, as it is one of the discordant areas. Third, having explained both the linear and nonlinear models for income taxation, we can build upon them to tackle more advanced issues – firstly tax avoidance and income shifting, followed by extensions on rent seeking and individual migration. Subsequently, empirical evidence on tax reforms and migration shall be also presented. Finally, a slight amendment to the hitherto prevailing utilitarian approach will be mapped.

5.1 Recent Debate between Mankiw and Diamond

The recent debate is well represented by two papers, starting with an overview from Mankiw et al. (2009). A response to this paper was then created by Diamond and Saez (2011). As both pieces of work adhere to a structure of several basic recommendations, they should be outlined below. The focus is paid to the six relevant areas. Mankiw et al. (2009) also summarize the development of actual tax systems with respect to their proposals, many of which were mentioned in section 2.2. Therefore, only the development not outlined earlier will be mentioned at this stage.

Marginal tax rate dependent on ability As Mankiw et al. (2009) claim, if the government could either observe ability directly or make citizens reveal their ability, the equity-efficiency trade-off could be avoided. The recommendation is that for the right balance in the equity-efficiency trade-off, the marginal tax rate

should depend on the distribution of ability. They conclude by calling this lesson too vague, therefore not of direct help to policy makers. To explain the vagueness, the first reason is the difficulty to observe ability, at least directly. This issue is discussed further in section 5.2. Secondly, a considerably different result is obtained dependent on the distribution of earnings used, lognormal or Pareto, which is more developed in the next paragraph. Diamond and Saez (2011) agree with this recommendation.

Marginal tax rate at high income This part is connected to the famous zero top marginal tax rate result, the logic behind which was already presented above in section 2.5. Mankiw et al. (2009) realize that this applies only to a single top earner. Without a top earner, the result does not apply. To explain, the basic model works with the strong assumption that the top earner is known beforehand, thus isolated from the rest through this rate. However, in the real world, there is naturally uncertainty who the top earner will be. Yet, Mankiw et al. (2009) still claim that this result might be instructive for policy makers. Concerning the result, it is crucial which distribution for income we use. The two disputed distributions are lognormal and Pareto. The main difference is that the Pareto distribution has its right tail thicker. This implies significant differences in tax rates for extreme values. Using the Pareto distribution then proclaims higher marginal tax rates optimal. These differences could be resolved by collecting more precise data on the distribution of earners, naturally not only in the USA, but in more countries. Another determinant is the social welfare function used, which will be reviewed in section 5.8. The most commonly used utilitarian approach suggests that the more inequality is present, the higher marginal taxes as a result. Thirdly, elasticities also play a role - as suggested by both linear and non-linear formulas for the optimal tax, the higher the elasticity of high earners, the lower the optimal marginal tax rate. Mankiw et al. (2009) conclude that under the assumptions of early work, the marginal tax rate at high income could be declining. Diamond and Saez (2011) expressed major disagreement with this suggestion. They claim that the Pareto distribution reflects the reality significantly closer than the lognormal distribution used by Mankiw et al. (2009). They furthermore claim that the test of distributions presented by

Mankiw et al. (2009) lacks power because of the data used and imprecise density fits. They therefore propose using formula (3) with $g = 0$ such that $\tau = 1/(1+a\cdot e)$, highlighting that this formula also does not require the assumption on homogeneity of preferences.[24] Last, but not least, one should also realize the the government might try to estimate the earnings level from which the zero (or low) marginal tax rate should be applied. If it does so through an estimation of some economic indicators of the taxed subjects where the highest values would be of decisive power, such highest value subjects might try to artificially lower such indicators to gain more favorable treatment. Of course, the government can expect such behavior and account for it in its estimation. Therefore, the top rate would be endogenous in such a case, implying a high complexity of the issue.

Optimality of a flat tax with a lump-sum transfer Mankiw et al. (2009) again present the optimality of a flat tax with a lump-sum transfer as crucially dependent on the shape of the distribution of ability. They claim that if the recommendations were derived from a wage distribution between lognormal and Pareto distributions, then it might be an optimal solution. Diamond and Saez (2011) reject this suggestion, highlighting the importance of the participation margin. They propose raising marginal tax rates due to declining marginal welfare weights and the Pareto shape of income distribution.

Inequality and redistribution Mankiw et al. (2009) claim that increased earnings potential of the top justifies more redistribution, which then slows the increase of the gap in disposable income. Diamond and Saez (2011) agree with this statement.

Tax system reflecting personal characteristics Mankiw et al. (2009) summarize that the theory suggests that any exogenous personal characteristic correlated with ability or preferences should be incorporated in the tax system. They claim that under the optimal tax model, the benefits of using gender and height as tools in the tax system would outweigh the administrative costs. This argument was

[24]While the standard Mirrlees (1971) model works with the assumption of preferences, modern literature strives to works without this assumption, as shown in for instance section 4.2.1.

further notably developed by Mankiw and Weinzierl (2010). Diamond and Saez (2011) present some disagreement, mainly highlighting the importance of the complexity and social acceptability of such a proposal. This issue will be put under scrutiny in section 5.2.

Capital income taxation Mankiw et al. (2009) argue that as supply of capital is highly elastic, the distortions to inter-temporal consumption and savings caused by capital taxation are too large to justify it. They propose that capital income should not be taxed, at least in expectation. Diamond and Saez (2011) strongly disagree, claiming that the argument used by Mankiw et al. (2009) does not imply that capital income should not be taxed.

5.2 Tagging

The notion of tagging, the use of personal characteristics for income taxation, was first noted already in the fundamental work of Mirrlees (1971). In fact, it was no later than in the fifth sentence of his contribution that tagging was mentioned. Short after, Akerlof (1978) further developed the idea and termed such variables "tags", working with the utilitarian framework which became central in further work. The main reason why tagging should be explored in depth here is that it has recently received considerable attention from economists. As Weinzierl (2012) points out, a clear advantage of tagging under the conventional theory is its achievement of redistribution without distorting work incentives, thus avoiding the classic equality-efficiency trade-off. It has been argued by Atkinson and Stiglitz (1980) that the main motive for the government not to use non-income characteristics are horizontal equity concerns.[25] In support of tagging Mankiw and Weinzierl (2010) pin not reflecting horizontal equity concerns as a fallacy of the standard approach, proposing more individual characteristics to determine taxation.

Piketty and Saez (2012) conclude that the two key concerns are the difficulty of the tax system and fairness. On the difficulty, it is believed that a tax system should be comprehensible. Therefore the fear is that the use of many tags would result in the

[25]The notion that equals should be treated equally (Weinzierl, 2012).

opposite. Concerning fairness, the goal is to understand how the government could use only a specific subset of taxes and ignore other tools, even taxes based on non-income characteristics correlated with ability, to maximize the social welfare. In reality, taxes and transfers already depend on certain characteristics, most notably family structure (number of dependent children), or disability status (permanent or temporary).

The work of Mankiw and Weinzierl (2010), who focus on taxation of height, will be shortly explained. Nevertheless, it remains paramount to realize that several other tags have been covered by scholars, most notably age, gender, and race (see Weinzierl, 2011; Alesina, Ichino, & Karabarbounis, 2011; Blumkin, Margalioth, & Sadka, 2009). Numerous other options for tags exist, the aim is that they are systematically related to ability and exogenous to individuals (Weinzierl, 2012).

Of course, discrimination also looms around the subject. One way to work around this issue might be to present rather a tax cut for the less able rather than an increase for the more able. However, it is not difficult to imagine that such positive discrimination would also be a target for criticism. Piketty and Saez (2012) argue that there is a more serious reason for governments to use actual earnings rather than proxies for ability in tax systems. One aspect they point out is mutability of the characteristic which would represent another obstacle to the implementation of tagging. To comment on the real situation, the characteristics typically used in tax and transfer system seem mostly to represent the need for support. Last, but not least, there are normative criteria behind utilitarianism which would explain the social unacceptability of such measures, even though it might cause failing to reach maximal welfare. These are further described in section 5.8.

5.2.1 Taxation of Height

Mankiw and Weinzierl (2010) base their study, which admits being utilitarian in its very name, on two grounds - theoretical and empirical. The theory as outlined above, states that an exogenous criterion correlated with ability should be used in the tax system. The empirical leg relies on three studies (see Judge & Cable, 2004; Persico, Postlewaite, & Silverman, 2004; Case & Paxson, 2008) which can be

summarized by stating that every additional inch of height brings an increase in earnings of about 1.5%. Therefore, the approach of Mankiw and Weinzierl (2010) is higher taxation of taller citizens.

In the theoretical model, they split the population into height groups. Within height groups, individuals are split into income brackets. They point out that in their model, height as a variable would improve the system, as excluding height as a variable again would bring us at the initial system. Therefore, this additional variable could not exacerbate the initial solution. To comment on the analytical results, they adhere to the zero marginal tax result extended to top earners among all height groups.

In the calculations, they choose to focus only on adult white males for simplicity who are split into three groups - tall, medium, short. Moreover, the age group considered is 32 to 39 in 1996 to control for the possibility of height trending over time.[26] They arrive at the average taxes higher by 16 percent for the tall than for the short.[27] To comment on the results, they obtain the average tax on the tall of 7.1% of the average tall income and on the other hand an average transfer of 13% to the short.[28] They give the example of a tall earner of income of $50,000 who would pay $4,500 more in taxes than a short earner with identical income. Last, they claim that such a tax policy would increase the welfare of the society by 0.19 percent of aggregate income of the economy. In addition, they note that the transfers to the short increase with risk aversion (CRRA) and the elasticity of labor supply, thus making the height tax more powerful to increase welfare.

In their conclusion, they explain that one of the reasons why height taxation could be accepted by the society is that it is not related to history of discrimination, unlike

[26]The sample size they in the end operate with includes 1,738 observations. They claim that adding women to the sample would likely increase the value of the height tax, as it would also serve as a proxy for gender-based taxes.

[27]There are two main explanations provided in the text. Firstly, adolescent height could cause characteristics valued in the labor market, such as self-esteem (Persico et al., 2004). Secondly, height could be correlated with productive attributes, such as cognitive ability (Case & Paxson, 2008). However, the source of the height pay premium is proclaimed insignificant for the problem of optimal height taxes.

[28]The elasticity of labor supply used for computations was 0.5, estimated by Piketty and Saez (2012) as high, but realistic.

race or gender. They further claim that genes could eventually assume the role of the ideal tag. Aware of the potential controversy, they point out that many might recoil from the idea of a height tax. They propose several explanations for such a reaction. Firstly, the utilitarian model might not reflect all constraints for political economy. The second explanation provided is the failure of the utilitarian model to account for horizontal equity concerns. They recognize the potential difficulties to explain to the tall why they pay a higher tax. Thirdly, the proposal to replace the utilitarian framework is given, giving the example of libertarian beliefs. In the end, they ask the reader to make a choice between advertising the height tax or rejecting, or significantly amending, the standard utilitarian approach.

To summarize, using additional personal characteristics seems to be a free lunch for tax systems, at least purely from the utilitarian point of view. A different matter is to what extent the utilitarian approach can be used to deliver actual policy recommendations. An alternative adjustment of the utilitarian approach, which explains the actual use of tags, is to be examined in section 5.8.

5.3 Tax Avoidance

It has been shown empirically by e.g. Saez et al. (2012) that tax avoidance can result from increases of tax rates. Tax avoidance is defined by Piketty and Saez (2012) as changes in reported income due to changes in the form of compensation but not in the total level of compensation. These opportunities arise when earners can transfer part of their income into another form or time period which is taxed more favorably. Examples of such behavior are transferring ordinary income into capital income, disguising vacation as business travel, or simply tax evasion using for instance off-shore accounts. Such changes might then result in changes in the size of the respective tax base.

At this point, it is key to realize the difference between real and tax avoidance responses. Real responses are determined by underlying individual preferences for work or consumption. On the other hand, tax avoidance responses are shaped solely by the tax system and the opportunities for tax avoidance it provides. It

is possible for policy makers to change the latter, but they can hardly change the real responses. For empirical purposes, it would be useful to divide elasticity to capture the different responses. One elasticity could still measure the real decline in economic activity, the second activity shifting. Piketty et al. (2011) focus on the topic of tax avoidance building upon the linear model and the main conclusions will be pointed out here.

First and most important, without any cost of enforcement, the optimal global policy to maximize tax revenue is to equalize taxes on all incomes. The elasticity in formula (2) would now represent the real elasticity. Moreover, the efficiency and ability to collect taxes can be also improved by equalizing the tax rates.[29]

Second, if the tax on one of the forms of income (for simplicity, let income be split into labor and capital) is zero, then the second form of income should be taxed to maximize revenue, i.e. $\tau^* = \frac{1}{1+e}$ as in equation (1).

Third, if both taxes are above zero, then a "fiscal externality" arises from sheltering,[30] as it causes shifts between sources of revenue. Then the tax on labor income should be above the revenue maximization rate. This is connected to the empirical discovery that virtually all short-term behavioral responses result from income shifting or re-timing, which generate fiscal externalities.

Fourth, two kinds of solutions of tax avoidance exist. The first kind represents those that can be fixed, even at no cost, for instance via the elimination of tax exempt local government bonds. The second sort are real enforcement constraints, for instance monitoring the informal business operating with solely cash transactions, or off-shore evasion.[31] The bottomline question is what fraction of tax avoidance elasticity can be eliminated by tax reforms.

To review, the problem of tax avoidance presents a challenge for tax collection. The

[29] Assuming that the poorly designed tax system, which could be fixed at no cost, is the only reason for avoidance opportunities.

[30] Sheltering can be described as legal ways of lowering taxable income, thus also tax liability.

[31] Off-shore evasion can be tackled more easily by several countries, as can currently be observed in the case of Cyprus or Switzerland.

main recommendation given is to close loopholes and equalize taxes on all sources of income. If discrepancies exist, fixing them can be either done at different costs, so it is crucial to estimate these before planning to curb them.

5.4 Income Shifting

To build on the previous section and the linear model, it should be noted that the government might have legitimate reasons to have different tax levels. Ramsey (1927) already recommended for the sake of efficiency to tax the most elastic goods or factors less than others. Continuing with the previously used labor and capital incomes for simplicity, the model can be extended to deliver four new results for revenue maximizing tax rates.[32] The two forms of income have different elasticities, thus enabling the same formula to deliver different numerical results.

To begin with, without income shifting, the respective rates are set as before, with the only difference resulting from a variation in the two rates due to the respective elasticities in formula (2).

Secondly, if income shifting opportunities exist, the two tax rates are brought closer. The mere existence of shifting opportunities does not necessarily decrease the ability to collect taxes, but only changes the relative mix of tax rates (Piketty & Saez, 2012). If shifting is infinite, the taxes are equalized and based on the average of the real elasticities.

Third, under the standard social welfare maximization objective, income distribution is another factor impacting optimal rates. The income which is more concentrated (with a concave social marginal utility of consumption) ought to be taxed more, ceteris paribus. Diamond and Mirrlees (1971) examined these distributive effects in optimal tax formulas under commodity taxation.

Summarizing, income shifting might not be a significant issue, if the tax rates reflect the different elasticities. Also, higher taxation of the more concentrated income is suggested.

[32]Other forms of income are individual and corporate earnings, or self-employment and employee earnings.

5.5 International Mobility

As was shown in part 2.1, international mobility is an aspect that ought to be taken into consideration due to the large share of total income among high earners. The European Union (EU) is a fine example, as labor mobility is one of its main targets and its costs have therefore decreased dramatically. While countries can usually implement policies treating immigration, emigration is usually not dealt with. The model of Mirrlees (1971) will be used to draw conclusions. Returning to the EU, it has become apparent that high earners are especially sensitive to taxes. There has been a debate on brain-drain, in the Czech Republic the case of doctors is well-known (Wiskow, 2006).[33] Mirrlees (1971) suggests that the higher the elasticity of migration, the lower the optimal tax. For empirical purposes, this elasticity can be regarded as another part of division of the overall elasticity, joining the real and income shifting elasticities. Naturally, the smaller the tax jurisdiction, the more limited its ability to impose higher taxes. The other way around, if we take the whole world as a jurisdiction, then elasticity of migration obviously becomes zero.

Policy makers have realized that the impact of the welfare effect of tax revenues and rates of neighboring jurisdictions might be significant, again especially for top earners. As a result, the claim appeared that tax coordination across jurisdictions could be mutually beneficial, also concerning loopholes. Therefore, endorsement to such development has been voiced both for individual (OECD, 2012a) and corporate (OECD, 2013) income taxes. Nonetheless, future action of individual countries and larger groups of countries is still difficult to predict. Yet, it would seem reasonable, especially in the times of fiscal hardships, to bring taxes closer and begin eliminating loopholes. One might claim that attempts to bring taxes closer together might be harmful for high-tax countries, as they would have to lower their tax rates, thus lowering their revenue. However, if consensus was struck, some lower limit on income taxation could be introduced to make international mobility less attractive, allowing high-tax countries to keep their rates. Some empirical evidence on international migration is to be presented in the next section.

[33]Though the doctors mostly demanded higher wages, taxes were not a concern in this case.

5.6 Empirical Evidence

Empirical evidence has been collected on both micro-level tax reforms and international migration. Saez et al. (2012) provide a survey of studies which compare changes in pre-tax incomes of groups affected by a tax reform to groups unaffected by it. This results in scrutiny of mainly short-term responses (1-5 years). Considerable variance in estimated elasticities exists, from below .25 to substantially larger estimates. Nevertheless, the large behavioral responses always result from tax avoidance in the form of retiming or income shifting. Consequently, large tax avoidance responses always result from poorly designed tax systems with arbitrage opportunities. Piketty and Saez (2012) point out that none of the empirical tax reform studies have shown large responses due to changes in real economic behavior.

To comment on international mobility, taxation differences within countries tend to be rather small, not much literature on the topic of mobility of individuals exists.[34] On the international level, capital income mobility has been explored for over 20 years.[35] Nonetheless, the field of international spatial mobility, where both tax differentials and mobility costs are considerably higher, has yet to be addressed by scholars. Some European countries have introduced preferential tax schemes for treatment of foreigners, for instance Denmark or France. This has indeed received attention of literature, the main works are by Kleven, Landais, and Saez (2010), Kleven, Landais, Saez, and Schultz (2013a), and Atkinson and Leigh (2010).

Kleven et al. (2010) explored the mobility of professional footballers in Europe. The conclusions reached were that the elasticity of mobility of domestic players with respect to the domestic net-of-tax rate is relatively small (around .15). The sensitivity of international players is suggested to be entirely different, allegedly approaching 1. The difference probably arises from the fact that most players play in their home country. Kleven et al. (2013a) extend the result to the case of international workers, using the Danish example of preferential taxes for foreigners,

[34]For an exploration of mobility within Swiss Cantons, see Kirchgassner and Pommerehne (1996).

[35]See e.g. Gordon, 1992 for a comparison of theoretical and actual practice, or Zodrow (2006) for an examination of capital income taxation in a small open economy.

reaching an elasticity of the number of foreign workers above one. These two results imply that the elasticity would likely be below .25, as the number of foreigners is relatively small (Piketty & Saez, 2012). As a result, the effect of international migration is unlikely to drastically effect the top tax rate. Having said that, taking into account the growing labor mobility and integration of labor markets, the effect might become larger as Piketty and Saez (2012) posit. They also claim that is advantageous from the perspective of a single country to introduce preferential tax schemes (as was the case in Europe recently), implementing beggar-thy-neighbor policies. Therefore, international regulation (e.g. by the EU) could prove itself beneficial.

Another part of empirical findings focused on cross-country and time data. The conclusion from studies by Atkinson and Leigh (2010), Roine et al. (2009), and Piketty et al. (2011) is that there is a strong negative correlation between top tax rates and top income shares (e.g. the top 1% of the income distribution). This correlation has allegedly been present both within and across countries. It is further claimed that the correlation is not mainly caused by income shifting responses. The key question then becomes whether it is caused by real supply side responses or rent-seeking effects (trickle-up). Piketty et al. (2011) suggest that rent-seeking effects play a role, implying that the top marginal tax rates could be substantially higher than is assumed, i.e. above 50-60%. Yet, Piketty and Saez (2012) call for more compelling empirical evidence to resolve this major question in optimal taxation of top earners.

To conclude, while international mobility is a concern for countries, empirical studies do imply that elasticity of foreign workers is below .25 due to the relatively low number of foreign employees. Yet, preferential tax schemes for foreigners could be implemented by countries where the cost of labor mobility is rather low, for instance in the EU.[36] More empirical evidence would be useful to reach a consensus on this significant concern.

[36]It can be claimed that such policies would be a step back to the *beggar-thy-neighbor* policies. Such policies were mainly used during the Great Depression in early 1930s and economists seem to be rather reluctant to such measures now.

5.7 Rent Seeking Effects

This part considers the case where the marginal economic product does not equal pay. This might be the case when executives are overpaid if in great position of authority, having the power to influence their own compensation (see e.g. Bebchuk & Fried, 2004 for an overview). A short summary of findings of Piketty et al. (2011), who extended the nonlinear model, shall be presented here. Nevertheless, it should also be noted beforehand that it is also related to the optimal tax rate with externalities in the form of charitable giving responses Saez (2004) or transfers across agents (Chetty, 2008a). To proceed to the main model, it delivers two major results. The first case is called *trickle-up* wherein top earners are overpaid with respect to their productivity. Then the optimal tax rate on them should be higher than $1/(1 - a \cdot e)$ from formula (3) with $g = 0$. Thus the name, as a tax cut on the top earners shifts economic resources towards the top from the bottom. In the extreme case where all behavioral responses are due to rent-seeking effects, $\tau = 1$. The second case called *trickle-down* considers the case of underpaid top earners with respect to their productivity. Conversely, the optimal tax τ becomes lower than in formula (3) with $g = 0$. Here, a tax cut on top earners shifts economic resources towards the bottom, as work of top earners benefits lower incomes, due to the underpay of top earners. On a general note, implementing an extended model capturing rent seeking effects also incorporates real economic product responses to tax changes. Such a model therefore becomes considerably more difficult to implement.

In a nutshell, if pay does not equal marginal product, the optimal tax rises if the pay is higher than the product. In the opposite case, the tax decreases (both compared to formula (3) with $g = 0$).

5.8 Revisiting Welfare Functions

As the welfare optima we obtain depend crucially on the model, tax rates, and possibly tags used, it is also useful to see what the underlying determinants of the model would be. As we saw in Table 1, the optimal tax rate τ varies sig-

nificantly depending on the criterion used. Several models have been mentioned throughout the text, namely the utilitarian, Rawlsian and median voter models. The utilitarian approach to taxation remains the main focus of the literature. The Pareto-efficiency of taxation has also been examined, suggesting that most general lessons from the utilitarian framework are kept (see e.g. Stiglitz, 1987; Werning, 2007). Another view to take into consideration is horizontal equity which can explain certain bounds in for instance progressivity or tagging. Also for this reason, Mankiw and Weinzierl (2010) called for an amendment to the standard utilitarian approach. Recently, an alternative approach has been suggested by Weinzierl (2012) where he attempts to bring together utilitarianism and libertarianism. As this piece of work sheds lights on horizontal equity and thus on both tagging and progressivity, a short summary of the work is to be presented.

As noted above, Weinzierl (2012) partly explains his motivation by the discrepancy of the conventional theory recommending tagging due to the decreased equity-efficiency trade-off on one hand, and the limited use of tagging in the real world. He claims that the answer is fundamental - the unrealistically narrow criterion of conventional optimal tax theory, utilitarianism. He claims that extensive evidence has proven that significant heterogeneity exists among individuals what criteria should be used for the design of the tax system, followed by stating that most people consider several criteria compelling. He aims to incorporate a normative perspective, which would include alternative criteria, into the already existing optimal tax theory, as the Mirrlees (1971) is considered fundamentally convincing. The elementary idea of the paper is augmenting the standard utilitarian approach by implementing the notion of equal sacrifice. Under equal sacrifice, all taxpayers should, as a result of taxation, face the same absolute reduction in their well-being from a laissez-faire starting point. Therefore, the recommendations of Mankiw and Weinzierl (2010) do not comply with this idea, as tall people face higher reductions in their well-being. As a result, under a scenario considering equal sacrifice, a tag is accepted if the resulting increase in aggregate utility outweighs the costs resulting from violating equal sacrifice.

Weinzierl (2012) puts forward three advantages of implementing the equal sacrifice criterion. Firstly, the discrepancy is explained between the theoretical recommendations on using tagging or tax progressivity on one hand, and the reality on the other. Secondly, the fraction of society which would identify itself with such an augmented approach would be larger due to its closer link to libertarianism. Third and related, it would better incorporate the indecisiveness of individuals concerning a single optimal approach. The author explains such an improvement via normal ambivalence of a majority of individuals towards redistribution.[37]

Weinzierl (2012) further points out that already Akerlof (1978) recognized the violation of horizontal equity as a drawback of tagging. Yet, he claims that horizontal equity is not a satisfactory explanation of limited tagging.[38] Conversely, he claims, equal sacrifice is more comprehensive with firm normative basis, and also happens to limit tagging. He further shows results of numerical simulations where the welfare gains of tagging on height, gender and race are substantial under the classic scenario, but neglectable if incorporating equal sacrifice. Nonetheless, if blindness is used as a tag, the effect of tagging remains sizable with equal sacrifice.[39] On tax progressivity, the author uses his theory to explain why the actual marginal tax rates on top earners are below the rates proposed by Diamond and Saez (2011).

In the conclusion, he states that the normative perspective has yet to be adopted by optimal tax research. The lesson proposed is that the lower the correlation of a tag to ability, the grater costs of the tag. He claims that the equal sacrifice criterion explains the two main aspects of the US tax and transfer system (limited tagging and considerable redistribution via tax progressivity). Moreover, the model suggests lower redistribution and lower marginal tax rates at high incomes, again

[37]The author notes that even the predisposed toward redistribution find compelling such normative principles that avoid the Rawls' concerns, as demonstrated by research by for instance Feldman and Zaller (1992).

[38]Due to the strong assumption of high administrative costs of tagging and the arbitrary choice of "horizontal" characteristics.

[39]Precisely and according to Weinzierl (2012), with the three tags of height, gender and race, the only two groups which have extra tax or transfer rate close to the one recommended for blindness (about 20%) are medium and short, nonwhite females. As there are together twelve strata in the first case, it seems reasonable to reject tagging to improve the well-being of two groups for most likely high administrative costs.

meeting reality, answering why such discrepancies exist.

To summarize, despite the development of the theory on optimal income taxation, complex models and much devotion to deriving optima, there is an underlying issue to solve first. The normative criteria used for the models are yet to be defined to determine the best policies. While many approaches incorporate numerous advantages, an approach which would capture at least a significant portion of the society is yet to be developed. The recent model of equal sacrifice by Weinzierl (2012) might be another step towards such a goal.

6 Conclusion

To conclude, the aim of this work was to summarize the topic of optimal taxation of top earners. First, the paper gave background on the topic. It is clear that the topic of income taxation is of utter importance, as it was shown that income taxation usually represents from about a third to one half of national income. Secondly, the trend of income concentration is projected to continue, showing the importance of taking the right approach towards top earners. Concerning the development of tax rates, we saw an increase from the beginning of the last century until about the end of World War II. After that, taxes have been overall on a decline, yet not reaching the low levels from the beginning of the last century. The number of tax brackets has also decreased, perhaps signaling a simplification of the tax systems. Next, it was explained that the goal of policy makers, at least in theory, is to maximize the welfare of the society, having raised enough revenue to finance the government. Moreover, two labor supply responses were explained – the intensive and extensive margins. As the last section of the background, theoretical development in the field was described to give a complex picture starting almost a century ago.

The first model explained was the linear model. Despite its limited tools compared to the nonlinear model, it is useful as a tool for becoming familiar with the logic behind both models. Both models were delivered bearing in mind three general guidelines: empirical relevance of economic mechanisms, robustness of results to changes in modeling assumptions, and feasibility of policy recommendations. The derivation of the optimal rate firstly showed that policy makers should strive to have taxes lower than the revenue maximizing rate. Both the revenue maximizing rate (formula 1) and the socially optimal rate (formula 2) were derived, the latter also including social weight. It was also shown that the recommendations of the linear model would meet real tax rates, especially those of the median voter.

The second model given was the more complex nonlinear model. After deriving formula (3) for the optimal top tax rate, the optimal tax rates at any income were given also for this case in formula (4). One should also bear in mind that even though continuous models were used for the optimal formulas, the new discrete

models also provide results close to those of the continuous models.

Another section mapped the discussion and more advanced issues concerning the topic of optimal taxation. Firstly, debate between the two main groups represented by N. Gregory Mankiw and Peter A. Diamond was presented. The areas that they agree upon are firstly that marginal tax rate should depend on ability, secondly that higher inequality implies more redistribution. On the disagreement, the two main issues are the progressivity of marginal tax rates at the top and the distribution of earners. Next, tagging is discussed, showing that under the conventional approach, it would be beneficial. Yet, as it might be difficult for the society to accept tags, the practical value is questionable. Subsequently, tax avoidance and labor mobility is central, hinting that we might have to account for several elasticities among top earners. The global solution to these issues would be an identical tax rate to all. However, the feasibility of such a proposal can be questioned. Then, rent seeking among top executives is described, hinting that the tax rate on them might be higher than the formulas suggest. Finally, the section concludes discussing which normative framework should be used to deliver recommendations.

This thesis serves well as an overview of the field of optimal income taxation. It provides the essential background and derivation of the main models. Moreover, it summarizes the discussion among prominent economists to provide the reader with the latest development in the field. Therefore, the main contribution of the thesis is providing the reader with a well-rounded summary of the topic. Concerning the future development, the resolution to the debate could be achieved firstly through more precise data on the distribution of income, finally allowing to decide whether the Pareto or lognormal distribution represents the reality better. Secondly, identification of all the actors, from policy makers through companies to citizens, would be useful, but then again make the problem more complex. However, this cost might be worth paying to achieve better results. Third, further normative development might also provide a useful insight into actual policy making and implementability. It seems that its future both theoretical and empirical works will remain exciting to follow.

References

Abramitzky, R. (2013). *The Mystery of the Kibbutz: How Socialism Succeeded.* (Under contract with Princeton University Press)

Abramitzky, R., & Lavy, V. (2011, May). *How Responsive is Investment in Schooling to Changes in Redistribution Policies and in Returns* (Working Paper No. 17093). National Bureau of Economic Research, Inc. Retrieved from http://ideas.repec.org/p/nbr/nberwo/17093.html

Adema, W., Fron, P., & Ladaique, M. (2011, November). Is the European Welfare State Really More Expensive? Retrieved from /content/workingpaper/5kg2d2d4pbf0-en doi: http://dx.doi.org/10.1787/5kg2d2d4pbf0-en

Akerlof, G. A. (1978, March). The Economics of "Tagging" as Applied to the Optimal Income Tax, Welfare Programs, and Manpower Planning. *The American Economic Review*, *68*(1), 8-19. Retrieved from http://www.ssc.wisc.edu/~scholz/Teaching_742/Akerlof.pdf

Alesina, A., Ichino, A., & Karabarbounis, L. (2011, May). Gender-Based Taxation and the Division of Family Chores. *American Economic Journal: Economic Policy*, *3*(2), 1-40. Retrieved from http://ideas.repec.org/a/aea/aejpol/v3y2011i2p1-40.html

Alm, J., Dickert Conlin, S., & Whittington, L. (1999). Policy watch: The marriage penalty. *The Journal of Economic Perspectives*, *13*(3), 193–204.

Altshuler, R., Harris, B., & Toder, E. (2011). *Capital Income Taxation and Progressivity in a Global Economy* (Departmental Working Papers). Rutgers University, Department of Economics. Retrieved from http://EconPapers.repec.org/RePEc:rut:rutres:201122

Atkinson, A. B. (1996). *Public Economics in Action: The Basic Income/Flat Tax Proposal* (No. 9780198292166). Oxford University Press. Retrieved from http://ideas.repec.org/b/oxp/obooks/9780198292166.html

Atkinson, A. B., & Leigh, A. (2010, May). *The Distribution of Top Incomes in Five Anglo-Saxon Countries over the Twentieth Century* (IZA Discussion Papers No. 4937). Institute for the Study of Labor (IZA). Retrieved from http://ftp.iza.org/dp4937.pdf

Atkinson, A. B., Piketty, T., & Saez, E. (2011, March). Top Incomes in the Long Run of History. *Journal of Economic Literature*, *49*(1), 3-71. Retrieved from http://ideas.repec.org/a/aea/jeclit/v49y2011i1p3-71.html

Atkinson, A. B., & Stiglitz, J. E. (1976). The design of tax structure: Direct versus indirect taxation. *Journal of Public Economics*, *6*(1-2), 55-75. Retrieved from http://ideas.repec.org/a/eee/pubeco/v6y1976i1-2p55-75.html

Atkinson, A. B., & Stiglitz, J. E. (1980). *Lectures on public economics*. London: McGraw-Hill.

Barr, N. A. (1993). *The economics of the welfare state*. Stanford University Press.

Bebchuk, L., & Fried, J. (2004). *Pay without Performance: The Unfulfilled Promise of Executive Compensation*. Harvard University Press: Cambridge.

Best, M., & Kleven, H. (2012). *Optimal Income Taxation with Career Effects of Work Effort* (Tech. Rep.). London School of Economics and Political Science. Retrieved from http://personal.lse.ac.uk/BESTM/research/best-kleven_april2012.pdf

Blumkin, T., Margalioth, Y., & Sadka, E. (2009, October). Incorporating affirmative action into the welfare state. *Journal of Public Economics*, *93*(9-10), 1027-1035. Retrieved from http://econ.hevra.haifa.ac.il/~todd/seminars/papers08-09/affirmative_action_15_february_2009_final.pdf

Cancian, M., & Danziger, S. (Eds.). (2009). *Changing Poverty, Changing Policies*. Russell Sage Foundation.

Carnegy, H. (2013, January). France hints at lowering 75% tax rate. *Financial Times*. Retrieved from http://www.ft.com/intl/cms/s/0/a2bb4172-5811-11e2-90c6-00144feab49a.html#axzz2KCSgZrn9

Case, A., & Paxson, C. (2008, June). Stature and Status: Height, Ability, and Labor Market Outcomes. *Journal of Political Economy*, *116*(3), 499-532. Retrieved from http://ideas.repec.org/a/ucp/jpolec/v116y2008i3p499-532.html

Chetty, R. (2008a, March). *Is the Taxable Income Elasticity Sufficient to Calculate Deadweight Loss? The Implications of Evasion and Avoid-*

ance (Working Paper No. 13844). National Bureau of Economic Research, Inc. Retrieved from http://www.economics.harvard.edu/faculty/chetty/files/taxable_income_formula_aej.pdf

Chetty, R. (2008b, October). *Sufficient Statistics for Welfare Analysis: A Bridge Between Structural and Reduced-Form Methods* (Working Paper No. 14399). National Bureau of Economic Research. Retrieved from http://www.nber.org/papers/w14399

Coltrane, S. (2004, March). Research on household labor: Modeling and measuring the social embeddedness of routine family work. *Journal of Marriage and Family*, *62*(4), 1208–1233. Retrieved from http://onlinelibrary.wiley.com/doi/10.1111/j.1741-3737.2000.01208.x/abstract

Daneshkhu, S., & Carnegy, H. (2013, March 12). Hollande blow as senior executives join French exodus. *Financial Times Europe*(38182), 2.

Diamond, P. A. (1980, February). Income taxation with fixed hours of work. *Journal of Public Economics*, *13*(1), 101-110. Retrieved from http://ideas.repec.org/a/eee/pubeco/v13y1980i1p101-110.html

Diamond, P. A. (1998, March). Optimal Income Taxation: An Example with a U-Shaped Pattern of Optimal Marginal Tax Rates. *American Economic Review*, *88*(1), 83-95. Retrieved from http://ideas.repec.org/a/aea/aecrev/v88y1998i1p83-95.html

Diamond, P. A., & Mirrlees, J. A. (1971). Optimal Taxation and Public Production I: Production Efficiency and II: Tax Rules. *The American Economic Review*, *61*, 8-27, 261-278.

Diamond, P. A., & Saez, E. (2011). *The Case for a Progressive Tax: From Basic Research to Policy Recommendations* (CESifo Working Paper Series No. 3548). CESifo Group Munich. Retrieved from http://ideas.repec.org/p/ces/ceswps/_3548.html

Eaton, J., & Rosen, H. S. (1980, September). Optimal Redistributive Taxation and Uncertainty. *The Quarterly Journal of Economics*, *95*(2), 357-64. Retrieved from http://ideas.repec.org/a/tpr/qjecon/v95y1980i2p357-64.html

Edgeworth, F. Y. (1897, 3). The Pure Theory of Taxation. *The Economic Journal*,

7(25), 46-70.

Eurostat Press Office. (2012, May). *Taxation trends in the European Union.* Press Release. Retrieved from http://europa.eu/rapid/press-release_STAT-12-77_en.htm#PR_metaPressRelease_bottom

Feldman, S., & Zaller, J. (1992, February). The political culture of ambivalence: Ideological responses to the welfare state. *American Journal of Political Science*, *36*(1), 268–307.

Flora, P. (1983). *State, economy, and society in Western Europe 1815-1975 : a data handbook in two volumes* [Book]. St. James Press.

Gordon, R. H. (1992, July). Can Capital Income Taxes Survive in Open Economies? *Journal of Finance*, *47*(3), 1159-80. Retrieved from http://ideas.repec.org/a/bla/jfinan/v47y1992i3p1159-80.html

Harrison, A. (1979). The Structure and Reform of Direct Taxation, report of a committee chaired by Professor J. E. Meade. *Journal of Social Policy*, *8*, 109-111. Retrieved from http://journals.cambridge.org/article_S0047279400008424 doi: 10.1017/S0047279400008424

Judge, T. A., & Cable, D. M. (2004). The effect of physical height on workplace success and income: preliminary test of a theoretical model. *Journal of Applied Psychology*, *89*(3), 428.

Kaplow, L. (2004, April). *On the Undesirability of Commodity Taxation Even When Income Taxation is Not Optimal* (Working Paper No. 10407). National Bureau of Economic Research. Retrieved from http://www.nber.org/papers/w10407

Kirchgassner, G., & Pommerehne, W. W. (1996, June). Tax harmonization and tax competition in the European Union: Lessons from Switzerland. *Journal of Public Economics*, *60*(3), 351-371. Retrieved from http://www.idheap.ch/idheap.nsf/9a665b63d3a8260cc1256a0800365df0/51c8a6dfa44ff463c125716a00365932/$FILE/POMMEREHNE_1996_tax_competition.pdf

Kleven, H. J., Claus Thustrup Kreiner, & Saez, E. (2006, November). *The Optimal Income Taxation of Couples* (Working Paper No. 12685). National Bureau of

Economic Research. Retrieved from http://www.nber.org/papers/w12685

Kleven, H. J., Landais, C., & Saez, E. (2010, November). *Taxation and International Migration of Superstars: Evidence from the European Football Market* (Working Paper No. 16545). National Bureau of Economic Research. Retrieved from http://ideas.repec.org/p/nbr/nberwo/16545.html

Kleven, H. J., Landais, C., Saez, E., & Schultz, E. (2013a). *Migration and Wage Effects of Taxing Top Earners: Evidence from the Foreigners' Tax Scheme in Denmark* (Tech. Rep.). Berkeley Working Paper. Retrieved from http://elsa.berkeley.edu/~saez/kleven-landais-saez-schultzNBER13danishscheme.pdf

Kleven, H. J., Landais, C., Saez, E., & Schultz, E. A. (2013b, March). *Migration and Wage Effects of Taxing Top Earners: Evidence from the Foreigners' Tax Scheme in Denmark* (Working Paper No. 18885). National Bureau of Economic Research. Retrieved from http://www.nber.org/papers/w18885

Laffer, A. B. (2004, June). *The Laffer curve: Past, present, and future.* Executive Summary Backgrounder. Retrieved from http://news.heartland.org/sites/all/modules/custom/heartland_migration/files/pdfs/15245.pdf

Laroque, G. R. (2005, April). Indirect taxation is superfluous under separability and taste homogeneity: a simple proof. *Economics Letters*, *87*(1), 141-144. Retrieved from http://ideas.repec.org/a/eee/ecolet/v87y2005i1p141-144.html

Mankiw, N. G., & Weinzierl, M. (2010, February). The Optimal Taxation of Height: A Case Study of Utilitarian Income Redistribution. *American Economic Journal: Economic Policy*, *2*(1), 155-76. Retrieved from http://www.nber.org/papers/w14976.pdf

Mankiw, N. G., Weinzierl, M., & Yagan, D. (2009, June). *Optimal Taxation in Theory and Practice* (Working Paper No. 15071). National Bureau of Economic Research. Retrieved from http://www.nber.org/papers/w15071

Milanovic, B. (2000). The median-voter hypothesis, income inequality, and income redistribution: an empirical test with the required data. *European*

Journal of Political Economy, *16*(3), 367 - 410. Retrieved from http://www.sciencedirect.com/science/article/pii/S0176268000000148 doi: 10.1016/S0176-2680(00)00014-8

Milligan, K., & Stabile, M. (2008, December). *Do Child Tax Benefits Affect the Wellbeing of Children? Evidence from Canadian Child Benefit Expansions* (Working Paper No. 14624). National Bureau of Economic Research. Retrieved from http://www.nber.org/papers/w14624

Mirrlees, J. A. (1971, April). An Exploration in the Theory of Optimum Income Taxation. *Review of Economic Studies*, *38*(114), 175-208. Retrieved from http://ideas.repec.org/a/bla/restud/v38y1971i114p175-208.html

Mirrlees, J. A. (2010). *Dimensions of Tax Design: The Mirrlees Review.* Oxford University Press, USA. Retrieved from http://www.amazon.com/Dimensions-Tax-Design-Mirrlees-Review/dp/0199553750%3FSubscriptionId%3D0JYN1NVW651KCA56C102%26tag%3Dtechkie-20%26linkCode%3Dxm2%26camp%3D2025%26creative%3D165953%26creativeASIN%3D0199553750

Mirrlees, J. A., & Stuart, A. (2011). *Tax By Design: The Mirrlees Review.* Oxford University Press, USA. Retrieved from http://www.amazon.com/Tax-By-Design-Mirrlees-Review/dp/0199553742%3FSubscriptionId%3D0JYN1NVW651KCA56C102%26tag%3Dtechkie-20%26linkCode%3Dxm2%26camp%3D2025%26creative%3D165953%26creativeASIN%3D0199553742

Mládek, J., & Havel, J. (2012). *Fair taxes and a balanced state budget!* (Party Program). ČSSD. Retrieved from http://www.cssd.cz/data/files/tax-policy.pdf

OECD. (2012a). *Hybrid Mismatch Arrangements: Tax Policy and Compliance Issues* (Tech. Rep.). OECD. Retrieved from http://www.oecd.org/ctp/aggressive/HYBRIDS_ENG_Final_October2012.pdf

OECD. (2012b, January). *OECD Tax Database.* Retrieved from http://www.oecd.org/tax/taxpolicyanalysis/oecdtaxdatabase.htm#C_CorporateCaptial

OECD. (2012c). *Revenue Statistics 1965-2012 / Statistiques des*

recettes publiques 1965-2012. Organization for Economic. Retrieved from http://www.keepeek.com/Digital-Asset-Management/oecd/taxation/revenue-statistics-2012_rev_stats-2012-en-fr

OECD. (2012d). *Special Feature: Trends in personal income tax and employee social security contribution schedules* (Tech. Rep.). OECD. Retrieved from http://www.oecd.org/tax/taxpolicyanalysis/50131824.pdf

OECD. (2013). *Addressing Base Erosion and Profit Shifting* (Tech. Rep.). OECD.

Persico, N., Postlewaite, A., & Silverman, D. (2004, May). *The Effect of Adolescent Experience on Labor Market Outcomes: The Case of Height* (Working Paper No. 10522). National Bureau of Economic Research, Inc. Retrieved from http://ideas.repec.org/p/nbr/nberwo/10522.html

Piketty, T. (1997). La redistribution fiscale face au chômage. *Revue française d'économie*, *12*(1), 157–201. Retrieved from http://www.persee.fr/web/revues/home/prescript/article/rfeco_0769-0479_1997_num_12_1_1016 doi: 10.3406/rfeco.1997.1016

Piketty, T., & Saez, E. (2001, September). *Income Inequality in the United States, 1913-1998 (series updated to 2000 available)* (Working Paper No. 8467). National Bureau of Economic Research. Retrieved from http://www.nber.org/papers/w8467

Piketty, T., & Saez, E. (2012, November). *Optimal Labor Income Taxation* (Working Paper No. 18521). National Bureau of Economic Research. Retrieved from http://www.nber.org/papers/w18521

Piketty, T., Saez, E., & Stantcheva, S. (2011, November). *Optimal Taxation of Top Labor Incomes: A Tale of Three Elasticities* (Working Paper No. 17616). National Bureau of Economic Research. Retrieved from http://www.nber.org/papers/w17616

Ramey, V. A., & Francis, N. (2009, July). A Century of Work and Leisure. *American Economic Journal: Macroeconomics*, *1*(2), 189-224. Retrieved from http://ideas.repec.org/a/aea/aejmac/v1y2009i2p189-224.html

Ramsey, F. P. (1927, March). A contribution to the Theory of Taxation. *The Economic Journal*, *37*, 47-61.

Research Institute for Labour and Social Affairs. (2012, June). *Main Economic and Social Indicators of the Czech Republic 1990 - 2011* (Tech. Rep.). Research Institute for Labour and Social Affairs.

Revelli, F. (2002). Local taxes, national politics and spatial interactions in English district election results. *European Journal of Political Economy*, *18*(2), 281 - 299. Retrieved from http://www.sciencedirect.com/science/article/pii/S0176268002000812 doi: 10.1016/S0176-2680(02)00081-2

Roháč, D. (2013, February). Where's Central Europe's Reform Zeal? *Central Europe Digest*. Retrieved from http://www.cepa.org/ced/view.aspx?record_id=373

Roine, J., Vlachos, J., & Waldenström, 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. Retrieved from http://www.sciencedirect.com/science/article/pii/S0047272709000383 doi: 10.1016/j.jpubeco.2009.04.003

Rosen, F. (1997). Utilitarianism and the Punishment of the Innocent: The Origins of a False Doctrine. *Utilitas*, *9*, 23–37. Retrieved from http://dx.doi.org/10.1017/S0953820800005112 doi: 10.1017/S0953820800005112

Sadka, E. (1976, June). On Income Distribution, Incentive Effects and Optimal Income Taxation. *Review of Economic Studies*, *43*(2), 261-67. Retrieved from http://ideas.repec.org/a/bla/restud/v43y1976i2p261-67.html

Saez, E. (1999). A Characterization of the Income Tax Schedule Minimizing Deadweight Burden (Doctoral dissertation, MIT). *unpublished PhD dissertation, MIT*.

Saez, E. (2001). Using Elasticities to Derive Optimal Income Tax Rates. *The Review of Economic Studies*, *68*(1), 205-229. Retrieved from http://restud.oxfordjournals.org/content/68/1/205.abstract doi: 10.1111/1467-937X.00166

Saez, E. (2002, August). Optimal Income Transfer Programs: Intensive Versus Extensive Labor Supply Responses. *The Quarterly Journal of Economics*, *117*(3), 1039-1073. Retrieved from http://ideas.repec.org/a/

tpr/qjecon/v117y2002i3p1039-1073.html

Saez, E. (2004, December). The optimal treatment of tax expenditures. *Journal of Public Economics*, *88*(12), 2657-2684. Retrieved from http://elsa.berkeley.edu/~saez/saezJpubE04charitable.pdf

Saez, E. (2013, 1). *Striking it Richer: The Evolution of Top Incomes in the United States* (Tech. Rep.). University of California, Berkeley. Retrieved from http://elsa.berkeley.edu/~saez/saez-UStopincomes-2011.pdf

Saez, E., Slemrod, J. B., & Giertz, S. H. (2012, May). The Elasticity of Taxable Income with Respect to Marginal Tax Rates: A Critical Review [Working Paper]. *Journal of Economic Li*, *L*(15012), 3-50. Retrieved from http://www.nber.org/papers/w15012

Seade, J. (1982, October). On the Sign of the Optimum Marginal Income Tax. *Review of Economic Studies*, *49*(4), 637-43. Retrieved from http://ideas.repec.org/a/bla/restud/v49y1982i4p637-43.html

Seade, J. K. (1977, April). On the shape of optimal tax schedules. *Journal of Public Economics*, *7*(2), 203-235. Retrieved from http://ideas.repec.org/a/eee/pubeco/v7y1977i2p203-235.html

Sheshinski, E. (1972). The Optimal Linear Income-Tax. *The Review of Economic Studies*, *39*, 279-302. Retrieved from http://darp.lse.ac.uk/papersdb/Sheshinski_(REStud_72).pdf

Stiglitz, J. E. (1982, August). *Self-Selection and Pareto Efficient Taxation* (Working Paper No. 632). National Bureau of Economic Research. Retrieved from http://www.nber.org/papers/w0632

Stiglitz, J. E. (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, p. 991-1042). Elsevier. Retrieved from http://ideas.repec.org/h/eee/pubchp/2-15.html

Varian, H. R. (1980, August). Redistributive taxation as social insurance. *Journal of Public Economics*, *14*(1), 49-68. Retrieved from http://ideas.repec.org/a/eee/pubeco/v14y1980i1p49-68.html

Vickrey, W. (1945, October). Measuring marginal utility by reactions to risk.

Econometrica: Journal of the Econometric Society, *13*, 319–333.

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* (HBS Working Papers No. 12-064). Harvard Business School. Retrieved from `http://www.hbs.edu/faculty/Publication%20Files/12-064.pdf`

Werning, I. (2007). Pareto efficient income taxation. *Massachusetts Institute of Technology, Manuscript*, *44*, 45.

Wiskow, C. (2006). *Health worker migration flows in Europe: overview and case studies in selected CEE countries-Romania, Czech Republic, Serbia and Croatia*. ILO.

Zodrow, G. (2006, May). Capital Mobility and Source-Based Taxation of Capital Income in Small Open Economies. *International Tax and Public Finance*, *13*(2), 269-294. Retrieved from `http://www.nek.lu.se/NEKAHA/Public%20finance/zodrow.pdf`